

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

09/540/245
Search for protein - protein search, using sw model

Run on: January 22, 2001, 11:50:43 ; Search time 233.01 Seconds
(without alignments)
223.791 Million cell updates/sec

Title: US-09-540-245A-2
Scarf score: 8316
Sequence: 1 MRGVGVNQMLSLSLGLVLAIL.....SSFVDEVEKVKVKGCTRCVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues ;

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8316	100.0	1525	20	Y17499	Human Slit-1 prote
2	8265	99.4	1529	20	Y27145	Human slit-2 prote
3	8265	99.4	1529	20	W96702	Full length slit-1
4	8137	97.8	1503	20	Y27142	Human slit-2 matur
5	8137	97.8	1503	20	W96701	Slit-like protein
6	8065	97.0	1529	21	Y76117	Rat Slit homologue
7	5717.5	68.8	1523	21	Y99395	Human PRO1336 (UNQ
8	5714.5	68.7	1523	20	Y27146	Human slit-3 prote
9	5714.5	68.7	1523	20	Y04137	Human slit 3 prote
10	5702.5	68.6	1496	20	Y27143	Human slit-3 matur
11	5702.5	68.6	1496	20	Y04136	Human slit 3 matur
12	5611.5	67.5	1523	20	Y14142	Human Slit protein

13	5597	67.3	1534	19	W46966	Amino acid sequenc
14	5589	67.2	1534	20	Y27144	Human slit-1 prote
15	5589	67.2	1534	20	Y04139	Human slit 1 prote
16	5589	67.2	1534	20	W96707	Protein sequence o
17	5588	67.2	1508	20	Y27141	Human slit-1 matur
18	5588	67.2	1508	20	Y04138	Human slit 1 matur
19	5588	67.2	1508	20	W96706	Protein sequence o
20	3543.5	42.6	716	21	Y76005	Rat Slit homologue
21	3443	41.4	1480	13	R25079	Drosophila SLIT pr
22	1668	20.1	299	20	Y27149	Protien encoded by
23	1668	20.1	299	20	W96703	EST clone protein
24	1429.5	17.2	450	20	Y27156	Peptide Seq ID No:
25	766	9.2	1010	20	W87896	Human JAGGED1 solu
26	766	9.2	1187	18	W18352	Proliferation and
27	766	9.2	1218	17	W05833	Human Serrate-1 (H
28	766	9.2	1218	19	W44301	Human serrate 1.
29	766	9.2	1218	20	W87894	Human JAGGED1 prot
30	766	9.2	1218	21	Y59597	Human Serrate prot
31	761	9.2	1036	18	W18351	Proliferation and
32	761	9.2	1218	18	W18354	Proliferation and
33	759	9.1	1208	19	W40827	Human Jagged prote
34	755	9.1	228	19	W46967	Amino acid sequenc
35	755	9.1	228	20	Y27147	Mouse slit protein
36	737.5	8.9	2471	20	Y06816	Human Notch2 (humN
37	736.5	8.9	1193	17	W05835	Chick Serrate. Ga
38	736.5	8.9	1193	21	Y59599	Chick Serrate prot
39	719.5	8.7	2321	19	W49698	Human Notch3 prote
40	697	8.4	1872	19	W68510	Partial human Notc
41	691.5	8.3	1964	20	W95557	Mus musculus notch
42	685	8.2	1148	20	W87895	Human JAGGED2 prot
43	680.5	8.2	1055	19	W44298	Human serrate 2 pr
44	680.5	8.2	1212	19	W44299	Human serrate 2.
45	680.5	8.2	1257	17	W05834	Human Serrate-2 (H

ALIGNMENTS

RESULT 1
Y17499
ID Y17499 standard; Protein; 1525 AA.
XX
AC Y17499;
XX
DT 04-AUG-1999 (first entry)
XX
DE Human Slit-1 protein.
XX
KW Human; Slit-1; Robo; modulation; identification; interaction.
XX
OS Homo sapiens.
XX
FN WO9925831-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24245.
XX
PR 07-APR-1998; 98US-0081057.
PR 14-NOV-1997; 97US-0065544.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Brose K, Goodman C, Kid T, Tessier-Lavigne M;
XX
DR WPI; 1999-347475/29.
DR N-PSDB; X76161.
XX
PT Human Slit polypeptide and related nucleic acids
XX
PS Disclosure; Page 19-21; 34pp; English.
XX
CC The present sequence is the human Slit-1 protein. The present invention

CC also describes a method for identifying agents which modulate the
CC interaction of Robo and a Robo ligand comprising: combining a Robo
CC polypeptide, a Slit polypeptide and a candidate agent under conditions
CC where the Robo and Slit polypeptides normally (but for the presence of
CC the agent) engage in a first interaction, where the Slit polypeptide
CC specifically binds, activated or inhibits the activation of the Robo
CC polypeptide and determining a second interaction of the Robo and Slit
CC polypeptides in the presence of the agent, where a difference between
CC the first and second interactions indicates that the agent modulates the
CC interaction of the Robo and Slit polypeptides; and a method to modulate
CC the interaction of Robo and a Robo ligand. The method is useful for
CC screening for Robo (roundabout) modulators and Robo:Slit complexes are
CC useful for regulating various cell functions, especially of neuronal
CC cells.

XX
SQ Sequence 1525 AA;

Query Match 100.0%; Score 8316; DB 20; Length 1525;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRGVGQMLSLSLGLVLAAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL 60
|
Db 1 mrgvgqmlslslglvllaailnkvpagpacqscsgstvdchglalrsvprniprnterl 60

Qy 61 DLGNNTIRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRNRNLQLFPE 120
|
Db 61 dlngnntiritktdfaglrhlrlvqlmenkistiergafqdlkelerlrnrnlqlfpe 120

Qy 121 LLFLGTAKLYRLDLSNQIAIPKAFRGAVDIKNQLDYNQISCIEDGAFRALRDLEVL 180
|
Db 121 llflgtaklyrldlsenqiaiprkafrgavdiqnldynqisciedgafalrdlevl 180

Qy 181 TLNNNNITRLSVASFNMHPKLRTRFRLHNNLYCDCHLAWLSDWLKRPRVGLYTQCMGPS 240
|
Db 181 tlannnitrslsvasfnnhmpklrtfrlhnnlycdchlawlsdwlkrprvglytqcmgps 240

Qy 241 HLRGHNVAEVQKREFVCSDEEEHGSFMAPSCSVLHCPAACTCSNNIVDCRGKLTETPT 300
|
Db 241 hlrghnvaevqkrefvcsdeeehgsfmapscsvlhcpaactcsnnivdcrgkltetpt 300

Qy 301 NLPETITEIRLEQNTIKVIPPAGFSPYKKLRRLDLSNNQISELAPDAFQGLRSLNSLVLY 360
|
Db 301 nlpetiteirleqntikvipagfspykkrrldlsnnqiselapdafqglrslnslvly 360

Qy 361 GNKITELPKSLFEGLSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLTQIAKGT 420
|
Db 361 gnkitelpkslfeglsqlllnankinclrldafqdlhnlslslydnkltqiakgtf 420

Qy 421 SPLRAIQTMHLAQNPFCIDCHLWLAADYLTNPISGARTSPRRANKRIGQIKSKKF 480
|
Db 421 splraiqtmhlaqnpficdchlwlaadyltnpietgartsprrankrigqikskkf 480

Qy 481 RCGSTEDYRSKLSGDCFADLACPEKRCCEGTVDSCSNQKLNKIPHIPOYTAE LRNNNE 540
|
Db 481 rcgstedyrsklsgcdcfadlacpekrccgtvdcscnqklnkipehipoytaelrnnne 540

Qy 541 FTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLENVQHMKFKG 600
|
Db 541 ftvleatgifkklpqlrkinfnnkitdiegafegasgvneilltsnrleqvqhmfkfg 600

Qy 601 LESLKTMLRSNRITCVGNSFGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLLAN 660
|
Db 601 leslktmlrsnritcvgnsfglssvrlslydnqittvapgafdtlhlslstlnllan 660

Qy 661 PFNCNCYLAWLGEWLKRKRITGNPRCKPYFLKEIPIQDVAIQDFTCDGNDNDSNCSPL 720
|
Db 661 pfncncylawlgewlkrkrivtgnprckpyflkeipiqdvaiqdfcdgndndnsnsp 720

Qy 721 SRCPTCTCLDTVVRCNKGKLVLPKGIPTDVTLYLDGNQFTLVFKELSNYKHLTLIDL 780
|
Db 721 srcptctcldtvvrcnkgkvlvpkgiptdvtlyldgnqftlvfkelsnykhltilidl 780

Qy 781 SNNRISTLSNQSFNSMTQLLTLILSYNRLRCIPPTFDGLKSLRLSLHGNDISVVPEGA 840
|
Db 781 snnristsnqsfnsmtqlltlilsynrlrcipptfdglksrlslshgndisvvpega 840

Qy 841 FNDLSALSHLAIGANPLYCDCNMQWLSWVSEYKEPGIARCAPGEMADKLLLTTPSKK 900
|
Db 841 fndlsalshlaiganplycdcnmqwlsdwkseykepgiarcagpgemadklllttppskk 900

Qy 901 FTCQGPVDVNLAKCNPCSNPCNDGTCNSDPVDFYRCTCPYFGKQDCDPIHACISN 960
|
Db 901 ftcqgpvdvnilakcnpcsnpcndgtcnsdpvdfyrctcpyfgkqdcdpihacisn 960

Qy 961 PCKHGGTCHLKEGEDGFWCICADGFECEGCEVNVDDCEDNDCENNSTCVDGINNYTCLC 1020
|
Db 961 pckhggchtchkegedgfwcicadgfecegecevnvddcedndcennstcvdginnytclc 1020

Qy 1021 PPETGELCEKLDLFCQADLNPCHQDSCILTPKGFCDCTPGYVGEHCDIDFDDCQDNK 1080
|
Db 1021 ppetgelceekldlfcqadlnpcqhdskiltpkgfcdctpgyvgehcdidfdcdqdnk 1080

Qy 1081 CKNGAHCITDAVNGYTCICPEGYSGLFCEFSPPMVLPRTPCNDNFCQGAQCIVRINEPI 1140
|
Db 1081 ckngahtcdavngytcicpegysglfcefppmvlprrtpcndnfcqgaqcivrinepi 1140

Qy 1141 CQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQNTITLQIATDEDSGILLYKGD 1200
|
Db 1141 cqclpgyggekeklsvvnfinkesylqipsakvrpqtntitlqiatdedsgillykgd 1200

Qy 1201 HIAVELYGRVRASYDTGSHPASAIYSVETINDGNFHVELLALDQSLSLSDGNGPKII 1260
|
Db 1201 hiavelygrvrasydtgshpasaiysvetindgnfhvellaldqslslsdgngpkii 1260

Qy 1261 TNLKQSTLNFDSPLYVGGMPGKSNVASLRQAPGNGTSFHGICIRNLYINSELQDFQKVP 1320
|
Db 1261 tnlkqstlnfdsplyvggmpgksnvaslrqapgngtsfhgicirnllyinselqdfqkvp 1320

Qy 1321 MQTGILPGCEPCHKKVCAHGTCPSSQAGFTCECEQGWGMPGLDQRTNDPCLNKCVHGT 1380
|
Db 1321 mqtgilpgcepchkkvcahgtcpssqagftceceqgwmpldqrtndpclnkcvhgt 1380

Qy 1381 CLPINAFSYSCKCLEHGGVLCDEEEDLFNCPQAIKCKHGCRLSGLQPYCECSSGYTG 1440
|
Db 1381 clpinafsysckclehggvvlcdeeedlfncpqaikckhgcrlsaglpycecssgytg 1440

Qy 1441 DSCDREISCRGERIRDYQKQGYAACQTTKVSRLRLECRGGCAGGQCCGLRSKRRKYSF 1500
|
Db 1441 dscdreiscrgerirdyqkqgyaacqttkvsrlrlecrggcaggqccglrskrrkysf 1500

Qy 1501 ECTDGSFVDEVEKVVKCGCTRCVS 1525
|
Db 1501 ectdgsfvdvekvvkcgctrcvs 1525

RESULT 2

Y27145

ID Y27145 standard; protein; 1529 AA.

XX

AC Y27145;

XX

DT 15-SEP-1999 (first entry)

XX

DE Human slit-2 protein (Seq ID No: 12 of JP11164690).

XX

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;
muscle; endocrine system.

XX

OS Homo sapiens.

XX

PN JP11164690-A.

XX

PD 22-JUN-1999.

XX

PF 05-DEC-1997; 97JP-0335435.
XX
PR 05-DEC-1997; 97JP-0335435.
XX
PA (ASAH) ASahi KASEI KOGYO KK.
XX
DR WPI; 1999-411830/35.
DR N-PSDB; X89162.
XX
PT New vertebrate slit protein - useful for diagnosis and treatment of
PT cancers in nerves, muscle and endocrine system
XX
PS Disclosure; Page 58-63; 102pp; Japanese.
XX
CC The invention relates to a vertebrate-derived protein containing an amino
CC acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein
CC has at least 55 % homology to one of sequences shown in Y27141-Y27143,
CC and has slit protease-like activity. The vertebrate slit proteins encoding
CC nucleic acid sequences have at least 60% homology to nucleic acid
CC sequences X89161-163. The vertebrate-derived proteins can be produced
CC recombinantly by transforming host cells with expression vectors
CC comprising the encoding nucleic acids. The proteins of the invention are
CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine
CC system.
XX
SQ Sequence 1529 AA;

Query Match 99.4%; Score 8265; DB 20; Length 1529;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 1 MRGVGQWQLSLGLVLAAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRTERL 60
Db 1 mrgvgwqlslslglvllailnkvpagpacqscscsgstvdchglalrsvprniprterl 60
Qy 61 DLNGNNITRITKTDFAGLRLRLVQLMENKISTIERGAFQDLKELERLRNLRNLQLFPE 120
Db 61 dlngnnitritktfdaglrhlrlvqlmenkistiergafqdlkelerlrlnrnlqlfpe 120
Qy 121 LLFLGTAKLYRLDLSENQIAIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL 180
Db 121 llflgtaklyrldlseqiaiprkafrgavdi knlqldynqisciedgafalrdlevl 180
Qy 181 TLNNNNITRLSVASFNMPLKRTFRLHNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPS 240
Db 181 tlnnnnitrlsvasfnhmpklrtfrlhnnlycdchlawlsdwlrrprvrglytqcmgps 240
Qy 241 HLRGHNAEVQKREFVCSDEEBGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 300
Db 241 hlrghnvaevqkrefvcs---ghgsfmapscsvlhcpaactcsnnivdcrgkglteipt 296
Qy 301 NLPETITEIRLEQNTIKVIPPGAFSPYKLRRLIDNNQISELAPDAFQGLRSLNSLVLY 360
Db 297 nlpetiteirleqntikvippgafspykllrridlnnqiselapdafqglrslnslvly 356
Qy 361 GNKITELPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLTIAKGT 420
Db 357 gnkitelpkslfegflslqllllnankinclr vda fqdlnlnlslydnkltiakgtf 416
Qy 421 SPLRAIQTMHLAQNPFICDCHLKLADLYHTNPISGARCTSPRLANKRIGQIKSKKF 480
Db 417 splraiqtmhlaqnpficdchklwladlyhtnpietgarctsprlankrigqikskkf 476
Qy 481 RCS-----GTEDYRSKLSGDCFADLACPEKRCCEGTTVDCSNQKLNKIPHIQYTA 532
Db 477 rcsakeqyfigtedyrsklsgcd fad lacpekrccegttvdcsnqklnkipehiqyta 536
Qy 533 ELRLNNEFTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLEN 592
Db 537 elrlnneftvleatgiffkklpqlrkinf snnkitdiegafegasgvneilltsnr len 596
Qy 593 VQHKMFKGLESKLTMLRSNRITCVGNDSFGLSSVRLSLYDNQITTVAPGAFDTLHSL 652

Db 597 vqhk mfk gles klt mlrs nrit cvgn ds figl ssvrllslydnqittvapgafdtl hsl 656
Qy 653 STLNLNLANPFCNCYLAWLGEWLKRRKRVITGNPRCQKPYFLKEIPQDVAIQDTCDDGN 712
Db 657 stlnlanpfcncylawlgewlrrkkrivtgnprcqkpyflkeipqdvaiqdtcdgn 716
Qy 713 DDNSCSPLSRCTECTCLDTVVRCNSKGLKVLPGKIPROVTLYLDGNGFTLVPKELSNY 772
Db 717 ddns csplsrctectcl dtvvrcns kglkvlpgkiprdvtelyldngftlvpkelsny 776
Qy 773 KHLTLIDLNNRISTLSNQSFNMTQLLTLLISYNRLRCIPPTFDGLKSLRLSLHGND 832
Db 777 khltlidl nsnristlsnqsf nmtql ltlisynrlrcipptfdglskrlslhngnd 836
Qy 833 ISVVEGAFNDLSALSHLAIGANPLYCDNCMWLSDVWVSEYKEPGIARCAPGEMADKL 892
Db 837 isvvegafndlsalshlaiganplycdncmwlsdvwvseykepgiarcagpgemadkl 896
Qy 893 LLTPSKKFTCQGPVDVNLAKCNPCLSNPCKNDGTCSNDPVDFYRCTCPYFGKQDCDV 952
Db 897 lltpskkftcggpvdvnilakcnpclsnpc kndgtcsndp dvyfrc tcpyfgkqdc dv 956
Qy 953 PIHACISNPCKHGGTCHLKEGEEDGFWICADGFEGENCEVNVDDCEDNDCENNSTCVDG 1012
Db 957 pihacisnpckhgg tchlkegeedgfwicad gfe gencev nvdcedndcennstcvdg 1016
Qy 1013 INNYTCLCPPEYTGELCEEKLDFAQDLNCPQHDSKICILTPKGFCDCTPGYVGEHCDID 1072
Db 1017 innytclcppeytgelceekldfcaqdlncpqhds kiciltpkgfcdctpgyvgehcdid 1076
Qy 1073 FDDCQDNCKNGAHCTDAVNGYTICPEGYSGLFCEFSPPMVLPRTPCDNFDQNGAQC 1132
Db 1077 fddcqdnckngahctdavngyticpegysglfcef sppmvlprtpcdnfdcqngaqc 1136
Qy 1133 IVRINEPICQLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGI 1192
Db 1137 ivrinepicqlpgyqgekceklvsvnfinkesylqip sakvrpqt niti lqiatdedsg i 1196
Qy 1193 LLYKGDKDHIAVELYGRVRASYDTGSHPSAISYVETINDGNFHVELLALDQSLSLSV 1252
Db 1197 llykgdkdhiavelygrvrasydtgshpsa isyvetindgnfhvellaldqslslsv 1256
Qy 1253 DGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSHFCIRNLYINSE 1312
Db 1257 dggnpkiitnlskqstlnfdsplyvggm pgsnvaslrqapqngtshfcir nlyinse 1316
Qy 1313 LQDFQKVPMTGILPGCEPCHKVCAHGTCPQSSAGFTCECQEGWGMPLCDQRTNDPCL 1372
Db 1317 lqdfqkvpmgtgilpgcepchkvcahgtcpqss agftcecg egwgmplcdqrtndpcl 1376
Qy 1373 GNKCVHGTCLPINAFFSYCKCLEGGHGVLCDEEEDLFNCPQAIKCKHKCRSLGQPYC 1432
Db 1377 gncvvhgtclpinaffsysckcleghgv lcadeed lfnpcqai kckhkcrlslgqpyc 1436
Qy 1433 ECSSGYTGDSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLR 1492
Db 1437 ecssgytgds cdreiscrgerirdyqkqgyaacqt tkkv srlecr ggcaggqccgplr 1496
Qy 1493 SKRRKYSPECTDGSFVDEVEKVVKCGCTRCVS 1525
Db 1497 skrrkyspectdgsf vdevekvvkcgctrcvs 1529

RESULT 3
W96702
ID W96702 standard; protein; 1529 AA.
XX
AC W96702;
XX
DT 15-APR-1999 (first entry)
XX
DE Full length slit-like protein sequence.

XX
KW Slit-like polypeptide; diagnosis; treatment; nervous disease;
KW thyroid disease; adrenal disease; muscular disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..26
FT /note= "signal peptide"
FT Protein 27..1529
FT /note= "mature protein, claimed in Claim 1"
XX
PN JP11018777-A.
XX
PD 26-JAN-1999.
XX
PF 09-JUL-1997; 97JP-0183683.
XX
PR 09-JUL-1997; 97JP-0183683.
XX
(ASAHI) ASAHI KASEI KOGYO KK.
DR WPI; 1999-161084/14.
DR N-PSDB; X14979.
XX
PT New slit-like polypeptide - useful for diagnosis and treatment of
PT nervous, thyroid, adrenal and muscular diseases
XX
PS Example 5; Page 27-31; 50pp; Japanese.
XX
CC The present sequence represents the full length sequence of a
CC slit-like polypeptide. The polypeptide is useful for the
CC diagnosis and the treatment of all nervous diseases, thyroid
CC diseases, adrenal diseases and muscular diseases.
XX
SQ Sequence 1529 AA;

Query Match 99.4%; Score 8265; DB 20; Length 1529;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 1 MRGVGWQMLSLSLGLVLAAILNKVAPQACPQCSCSGSTVDCHGLALRSVPRNIPRINTERL 60
Db 1 mrgvgwqmlslslglvllaailnkvapqacpqcscsgstvdchglalrsvprniprnterl 60

Qy 61 DLGNNTIRITKTDFAGLRHLRVLQLMENISTIERGAFQDLKELERLRNRLHLQFPPE 120
Db 61 dlngnntirittktdfaglrhlrvlqlmenistiergafqdlkelerlrlnrnhlqlfpe 120

Qy 121 LLPLGTAKLYRLDLSNQIAIPKAFRGAVDIKNLQLDYNQISCIEDGAFAIRLDLEVL 180
Db 121 llplgtaklyrldlsenqiaiprkafrgavdi knlqldynqisciedgafralrdlevl 180

Qy 181 TLNNNNITRLSVASFNMPLKRTFRLHSNNLYCDCHLAWLSDWLKRPVGLYTQCMGPS 240
Db 181 tlnnnnitrslvasfnhmpklrtfrlshnnlycdchlawlsdwlkrprvglytqcmgps 240

Qy 241 HLRGHNAEVQKREFVCSDEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 300
Db 241 hlrghnaevqkrefvcs----ghqsfmapscsvlhcpaactcsnnivdcrgkglteipt 296

Qy 301 NLPETITEIRLEQNTIKVIPPAGAFSPYKKLRIDLSNNQISELAPDAFQGLRSLNSLVLY 360
Db 297 nlpetiteirleqntikvipagafspykkrridlsnnqiselapdafqglrslnslvly 356

Qy 361 GNKITELPKSLFEGLFSLQLLLLNNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAGTGF 420
Db 357 gnkitelpkslfegfslqllllnankinclrldafqdlhnl nllslydnklqtiagt 416

Qy 421 SPLRAITQTMHLAQNPFCIDCHLKLWADLYHTNPISGARTSPRRLANKRIGQIKSKKF 480
Db 417 splraitqtmhlaqnpficdchklwladlyhtnpietgarctsprlankrigqikskkf 476

Qy 481 RCS-----GTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPHEHIPQYTA 532
Db 477 rcsakegyfipgtedyrsklsqdcfadlacpekrcegttvdcsnqklnkipehipqyta 536

Qy 533 ELRLNNEFTVLEATGIFKKLPQLRKINFSNNKITDIEEAGFEGASGVNEILLTSNRLEN 592
Db 537 elrlnneftvleatgiffkklpqlrkinfnsnnkitdieeagfegasgvneilltsnrln 596

Qy 593 VQHKMFGLSLKTLMLRSNRITCVGNSFIGLSSVRLSLYDNQITTVAPGAFDTLHSL 652
Db 597 vqhkmfgleslktlmlrsnritcvgnsfiglssvrlslydnqittvapgafdtlhlsl 656

Qy 653 STLNLNANPFNCNCLAWLGEWLKRRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGDN 712
Db 657 stlnlanpfncncylawlgewlkrkrivtgnprcqkpyflkeipiqdvaigftcdgdn 716

Qy 713 DDNSCSPLSRCTECTCLDTVVRCNKGKLVLPKGIPTVTELYLDGNQFTLVPKLSNY 772
Db 717 ddnsccsplsrctectcldtvvrcsnkgklvlpkgiprtvteilyldgnqftlvpkelsny 776

Qy 773 KHLTLIDLNNRISTLSNQSFNSMTQLTLILSYNRLRCIPPRTFDGLKSLRLLSLHNGD 832
Db 777 khltlidlennristlsnqsfnsmtqltlilsynrlrcipprtfdglkslrllslhngd 836

Qy 833 ISVVPFEGAFNDLSALSHLAIGANPLYCDNCNMQWLSDWKSEYKEPGIARCAPGEMADKL 892
Db 837 isvvpega fndlsalshlaiganplycdncnmqwlswkseykepgiarcagpgemadkl 896

Qy 893 LLTPSPKFTCPQGPVDVNLAKCNPCLSNPKNDGTCNSDPVDFYRCTCPYFGKQDCDV 952
Db 897 lltpskftcpgpvdvnlakcnpclsnpkndgtcnsdpvdfyrctcpyfgkqdcdv 956

Qy 953 PIHACISNPKHGGTCHLKEGEEDGFWICADGFECEGCEVNVDDCEDNDCENNSTCVDG 1012
Db 957 pihacisnpckhgggtchlkegeedgfwicadgfecegecevnvddcedndcennstcvdg 1016

Qy 1013 INNYTCLCPPEYTGELCEEKLDFAQDLNCPQHDSKILTPKGFCDCTPGYVGEHCID 1072
Db 1017 innytclcppeytgelceekldfcaqdlncpqhdskciltpkgfcdctpgyvgehcid 1076

Qy 1073 FDDCQDNCKNGAHCTDAVNGYTICPEGYSGLFCEFSPPMVLPRTPSPCDNFCQNGAQ 1132
Db 1077 fddcqdnckngahctdavngyticpegysglfcefspmvlprrtpspcdnfcqngaq 1136

Qy 1133 IVRINEPICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGI 1192
Db 1137 ivrinepicqclpgyqgekecklvsvnfinkesylqipsakvrpqtntitlqiatdedsg 1196

Qy 1193 LLYKGDKHIAVELYRGRVRSYDTGSHPSAISYVETINDGNFHVHALLDQSLSLSV 1252
Db 1197 llykgdkhiavelyrgrvrsydtgshpsaisyvetindgnfhvhalldqslslsv 1256

Qy 1253 DGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPQNGTSFHGICRLNLYNSE 1312
Db 1257 dggnpkiitnlskqstlnfdsplyvggmpgksnvaslrqapqngtsfhgicrlnlyns 1316

Qy 1313 LQDFQKVPMTGILPGCEPCHKVCAHGTCPQSSQAGFTCECQEGWMPGLCDQRTNDPCL 1372
Db 1317 lqdfqkvpmgtgilpgcepchkvcahgtcpqssqagftcecqegwmpglcdqrtndpcl 1376

Qy 1373 GNKCVHGTCLPINAFFSYCKLEBHGVLDEEEDLFNCPQAIKCKHKGKRLSLGLQPPYC 1432
Db 1377 gncvghgtclpinaffsyscklebhgvldedeedlfncpqaikckhkgkrlslglqppyc 1436

Qy 1433 ECSSGYTGDSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLR 1492
Db 1437 ecssgytgdsdrcdreiscrgerirdyqkqgyaacqttkkvsrlecrggcaggqccgplr 1496

Qy 1493 SKRRKYSFECTDGSSFVDEVEKVVKCGCTRCVS 1525
Db 1497 skrrkysfectdgssfvdevekvkcgctrcvs 1529

Query Match 97.8%; Score 8137; DB 20; Length 1503;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1494; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy	27	ACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITRIKTKDFAGLRHLRVLQL	86
Db	1	acpaqcscsgstvdchglalrsvprniprnterldlngnnitritktdfaglrhlrvql	60
Qy	87	MENKISTIERGAFQDLKELERLRLNRNHQLFPPELLFGTAKLYRLDLSENQIQAIPRKA	146
Db	61	menkistiergafqdlkelerlrlnrnhqlfpellfigtaktlyrldlsenqiqaiprka	120
Qy	147	FRGAVIDKNLQLDYNQISICIEDGAFFRALRDLEVLTLN>NNNITRLSVASFNHMPKLRTFRL	206
Db	121	frgavidknlgldynqisciedgafralrdlevltlnnnnitrsvasfnhmpklrtfrrl	180
Qy	207	HSNNLYCDCHLAWSDLWRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDDEEGHQS	266
Db	181	hsnnlycdchlawsdwlrqprvglytqcmgpshlrghnvaevqkrefvc---ghqs	236
Qy	267	FMAPSCSVLHCIPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGAFSF	326
Db	237	fmapscsvlhcpaactcsnnivdcrgkglteiptnlpetiteirleqntikvippgafsf	296
Qy	327	YKKLRRLDISNNQISELPADAFOGLRSLNSLVLYGNKITTELPKSLFEGLSLOLLLLNAN	386

Db	297	ykklrridlsnnqiselpadafqglrslnsilvlygnkitelpksifegflsqllllnan	356
Qy	387	KINCLRVDAFDQLHNLNLLSLYDNKLQTIAGTFSPLRAIQTMHLAQNPFCDCHLKWL	446
Db	357	kinclrvdafqdlhnlsllydnklqtiagtfsplraiqtmhlaqnpfcdchklwla	416
Qy	447	DYLHTNPJETSGARCTSPRLANKRIGQIKSKKFRCS-----GTEDYRSKLSGDCFA	498
Db	417	dylhtnpjietsgarctsprrlankrigqikskkfrcsakegyfipgtderysklsgdcfa	476
Qy	499	DLACPEKCRCEGTTVDCSNQKLNKPIHEPIQYTAELRLNNEFTVLEATGIFKKLPQLRK	558
Db	477	dlacpekcrcegttvdcsnqklnkpihepipqytaelrlnneftvleatgifkklpqlrk	536
Qy	559	INFSNNKITDIEEGAFEGAGSGVNEILLTSNRLENVQHKMFKGLESKLTMLRSNRITCVG	618
Db	537	infsnnkitdieegafegagsgvneilltsnrlenvqhkmgfkgleskltmlrsnritlevg	596
Qy	619	NDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFNCNCLAWLGEWLRKK	678
Db	597	ndsfiglssvrlslydnqittvapgafdtlhlslstlnllanpfnccnclawlgewlrkk	656
Qy	679	RIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNNSCPLSRCPTECTCLDVTVVRCSN	738
Db	657	rivtgnprcqkpyflkeipiqdvaiqdfcdgndnnsclsrpcptectcldtvvrcsn	716
Qy	739	KGLKVLPGKIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQFSNMTQ	798
Db	717	kglkvlpgkiprdvtelyldgnqftlvpkelsnykhltilidlsnnristlsnqfsnmtq	776
Qy	799	LTLILSYNRLRCPPTFDGLKSLRLLSLHGNDISVVEGAFNDLSALSHLAIGANPLY	858
Db	777	ltililsynrlrcpprtfdglkslrlslhgndisvvegafndlsalshlaiganply	836
Qy	859	CDCNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTSPKKFTCQGPDVNVILAKNPC	918
Db	837	cdcnmqwlsdwvkseykepgiarcagpgemadkllltspkkftcggpvdvnilacnpc	896
Qy	919	LSNPCKNDGTCNSDPVDFYRCTCPYFGKGQDCDVPFHACISNPCKHGGTCHLKEGEEDGF	978
Db	897	lsnpckndgtcnsdpvdfyrctcpyfgkgqdcdpvhacisnpckhgggtchlkegeedgf	956
Qy	979	WCICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFAQ	1038
Db	957	wcicadgfevencevnnvddcedndcennstcvdginnytcclppeytgelceekldfaq	1016
Qy	1039	DLNQCQHDSCILTPKGFKDCTPGYVGEHCDIFDDCQDNKCKNGAHTDAVNGYTCIC	1098
Db	1017	dlnpcqhdsckiltpkgfkdcctpgyvgehcdifddcqdnkckngahtdavngytcic	1076
Qy	1099	PBGYSGLFCEFSPPMVLPRTPSPCDNFDQNGAQCIVRINEPICQLPGYQGEKCEKLVSV	1158
Db	1077	pegysglfcefesppmvlprtspcdnfdqngaqcivrinepicqlpgyqgekceklvsv	1136
Qy	1159	NFINKEYSYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIARELYRGRVRSYDTG	1218
Db	1137	nfinkeysylqipksakvrpqtntilqiatdedsgillykgdkdhiarelyrgrvrsydtg	1196
Qy	1219	SHPASAIYSVETINDGNFHIIVELLALDQSLSLSDVGGNPKIITNLKQSTLNFDSPLYVG	1278
Db	1197	shpasaiysvetindgnfhiivellaldqslslsdvggnpkiitnlkqstlnfdsplyvg	1256
Qy	1279	GMPGKSNVASLRQAPGNGTSFHGICRLNLYINSELQDFQKVPMTGILPGCEPCHKVCA	1338
Db	1257	gmpgksnvaslrqapgngtsfhgicrlnlyinselqdfqkvpmqtgilpgcepchkvca	1316
Qy	1339	HGTCQPSSQAFTCECQEGWGMPLCDQRTNDPCLGNKCVHGTCLPINALFSYSCKLEGHG	1398
Db	1317	hgtcqpssqaftceceqegwgmplcdqrtndpclgnkcvhgtclpinalfsysckleghg	1376
Qy	1399	GVLCDDEEDLNFPCQAIKCKHGKCRSLGSLGQPYCECSSGYTGSDCDREISCRGERIRDY	1458

Db 1377 gvlcodeedlnpcqgaickhkgkrlsglqgpycecsstgtdscdreiscrgerirdyy 1436

Qy 1459 QKQGYAACQTTKKVSRLECRGGCAGGCCGGLRSLRRKYSFECTDSSSFVDEVEKVVVK 1518

Db 1437 gkqggyaacgttkkvsrlecrggcaggccgplrrkrrkysfectdsssfvdevekvvk 1496

Qy 1519 GCTRCVS 1525

Db 1497 gctrcvs 1503

RESULT 5

W96701

ID W96701 standard; protein; 1503 AA.

XX

AC W96701;

XX

DT 15-APR-1999 (first entry)

XX

DE Slit-like protein amino acid sequence.

XX

KW Slit-like polypeptide; diagnosis; treatment; nervous disease;
thyroid disease; adrenal disease; muscular disease.

XX

OS Homo sapiens.

XX

PN JP11018777-A.

XX

PD 26-JAN-1999.

XX

PF 09-JUL-1997; 97JP-0183683.

XX

PR 09-JUL-1997; 97JP-0183683.

XX

PA (ASAH) ASahi KASEI KOGYO KK.

XX

DR WPI; 1999-161084/14.

XX

PT New slit-like polypeptide - useful for diagnosis and treatment of
nervous, thyroid, adrenal and muscular diseases

XX

PS Claim 1; Page 17-21; 50pp; Japanese.

XX

CC The present sequence represents the mature protein sequence of
a slit-like polypeptide. The polypeptide is useful for the
diagnosis and the treatment of all nervous diseases, thyroid
diseases, adrenal diseases and muscular diseases.

XX

Q Sequence 1503 AA;

Query Match 97.8%; Score 8137; DB 20; Length 1503;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1494; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 27 ACPAQCSGSGSTVDCHGLALRSVPRNIPRTERLDLNGNNITRITKTFAGLRHLRVQL 86

Db 1 acpaqcscsgstvdchglalrsvprniprnterldlgnnnitritktfdaglrhlrvql 60

Qy 87 MENKISTIERGAFQDLKELERLRLNRNLQFPPELLFLGTAKLYRLDLSNQIAIPKA 146

Db 61 menkistiergafqdlkelerlrlnrnlqlfpellflgtaklyrldlsenqiaipka 120

Qy 147 FRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLRFTFL 206

Db 121 frgavdi knlqldynqisciedgafalrdlevltl nnnnitrlsvasfnhmpklrtf rl 180

Qy 207 HSNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPHLRGHNVAEVQKREFVCSDEEBGHQS 266

Db 181 hsnlycdchlawlsdwlrrprvglytqcmgphlrgnhvaevqkrefvcs---ghqs 236

Qy 267 FMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPAGFASP 326

Db 237 fmapscsvlhcpaactcsnnivdcrgkglteiptnlpetiteirleqntikvippgafsp 296

Qy 327 YKLRRLDLSNNQISELAPDAFQGLRSLNSLVYGNKITELPKSLFEGLFSLQLLLNAN 386

Db 297 yklrrldlsnnqiselapdafqglrslnslvlygnkitelpkslfeqlfslqllln nan 356

Qy 387 KINCLRVDAFQDLHNLNLSLYDNKLTQIAKGTFSPLRAIQTMLHAQNPPICDCHLKWLA 446

Db 357 kinclrvdafqdlhnl nlslydnkltqiakgtfsp lraigt mhaqnpficdchlk wla 416

Qy 447 DYLTNPDIETSGARCTSPRLANKRIGQIKSKKFRCS-----GTEDYRSKLSGDCFA 498

Db 417 dyltnpietsgarctsprrlankrigqikskkfrcsakeqy fipgtedyrsklsgdcfa 476

Qy 499 DLACEKCRCEGTTVDCSNQKLNKIPHIQYTAELRLNNEFTVLEATGIFKKLPQLRK 558

Db 477 dlacekcrcegttvdcsnqklnkipehipqytaelrl nneftvleatg ifkklpqlrk 536

Qy 559 INFNNKITDIEEGAFEGASGVNEILLTSNRLNVQHKMFQGLSKLTMLRSNRITCVG 618

Db 537 infnnkitdieegafegasgvneilltsnrlnvqhk mfgleslktmlrsnritcv g 596

Qy 619 NDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLANPFCNCYLAWGEWLK 678

Db 597 ndsfiglssvrlsl ydnqittvapgafdtl hslstlnlanpfcncylawgewl rk 656

Qy 679 RIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNSCSPLSRCPTCTCLDTVVRC 738

Db 657 rivtgnprcqkpyflkeipiqdva iqdftcdgndnscsplsrcptctcl dtvvrcsn 716

Qy 739 KGLKVLPGKIPROVTLYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQSF 798

Db 717 kglkvlpkgiprdvtelyldgnqftlvpkelsnykhl tldlssnnristlsnqsf smtq 776

Qy 799 LTLILSYNRLRIPCPTFDGLKSLRLSLHGNDSVVEGAFNDLSALSHLAIGANPLY 858

Db 777 ltlilsynrlrcipptfdglsrlslhgn dsvvgegafndlsalshlaiganply 836

Qy 859 CDCNMQWLSDWVWKEYKEPGIARCAPGEMADKLLLTTPSKKFTCGQPVVDNLAKN 918

Db 837 cdcmqwlswvkwkeykepgiarcapgemadklllt tpskkftcgqpvvdnlakncp 896

Qy 919 LSNPCKNDGTCNSDPVDFYRCTCPYGFQGDQDVPPIHACISNPCKHGGTCHLKEGEEDGF 978

Db 897 lsnpcndgtcnsdpvdfyrctcp ygfqgdcdvp i hacisnpckhgg tchlkegeedgf 956

Qy 979 WCICADGFEGENCEVNVDDCEDNDCENNSTCVGINNYTCLCPPEYTGECEKLDCAQ 1038

Db 957 wcicadgfegeencevndcedndcennstcv ginnnytclcppeytgelceekldcaq 1016

Qy 1039 DLNPQCQHSKCIILTPKGFKCDCTPGYVGEHCDIDFDDCQDNCKNGAHTDAVNGYTCIC 1098

Db 1017 dlnpcqhsckciiltpkgfkcdctpgy vgehcdid fddcdqdnckngahctdavn gytic 1076

Qy 1099 PEGYSGLFCEFSPPMVLPRISPCDNFDCQNGAQCIVRINEPICQLPGYQGEKELVSV 1158

Db 1077 pegysglfcefspmvlp rtpspcdnfdcqngaqcivrinepicqlpgyqgekelvsv 1136

Qy 1159 NFINKESYLQIPSAKVRPQTNIQLIATDEDSGILLYKGDKDHIAVELYGRVRASYDTG 1218

Db 1137 nfinkesylqip sakvrpqtntlqliatdedsgillykgdkdhia velygrvrasydtg 1196

Qy 1219 SHPASAIYSVETINDGNFHVELLALDQSLSLSDVGGNPKIITNLKQSTLNFDSPLYVG 1278

Db 1197 shpasaiysvetindgnfhvella ldqslslsdvggnpk iitnlkqstlnf dsplyvg 1256

Qy 1279 GMPGKSNVASLRQAPGQNGTSFHGICRNLINSELQDFQKVPMTGILPGCEPCHKVCA 1338

Db 1257 gmpgksnvaslrqapgngtsfhgicrnl in selqdfqkvpm t g ilpgcepchkvca 1316

Qy 1339 HGTCQPSQAGFTCECQEGWMGPLCDQRTNDPCLGNKCVHGTCLPINAFYSCKCLEGHG 1398

Db 1317 hgtcqpsqagftcecgqegwmgplcdqrtndp clgnkcvhgtcl pinafysckcleghg 1376

[illegible]

```

RESULT 7
Y99395
ID Y99395 standard; Protein; 1523 AA.
XX
AC Y99395;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1336 (UNQ691) amino acid sequence SEQ ID NO:198.
XX
Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
XX 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.

```

PR	10-SEP-1998;	98US-0099792.
PR	10-SEP-1998;	98US-0099808.
PR	10-SEP-1998;	98US-0099812.
PR	10-SEP-1998;	98US-0099815.
PR	10-SEP-1998;	98US-0099816.
PR	15-SEP-1998;	98US-0100385.
PR	15-SEP-1998;	98US-0100388.
PR	15-SEP-1998;	98US-0100390.
PR	16-SEP-1998;	98US-0100584.
PR	16-SEP-1998;	98US-0100627.
PR	16-SEP-1998;	98US-0100661.
PR	16-SEP-1998;	98US-0100662.
PR	16-SEP-1998;	98US-0100664.
PR	17-SEP-1998;	98US-0100683.
PR	17-SEP-1998;	98US-0100684.
PR	17-SEP-1998;	98US-0100710.
PR	17-SEP-1998;	98US-0100711.
PR	17-SEP-1998;	98US-0100919.
PR	17-SEP-1998;	98US-0100930.
PR	18-SEP-1998;	98US-0100848.
PR	18-SEP-1998;	98US-0100849.
PR	18-SEP-1998;	98US-0101014.
PR	18-SEP-1998;	98US-0101068.
PR	18-SEP-1998;	98US-0101071.
PR	22-SEP-1998;	98US-0101279.
PR	23-SEP-1998;	98US-0101471.
PR	23-SEP-1998;	98US-0101472.
PR	23-SEP-1998;	98US-0101474.
PR	23-SEP-1998;	98US-0101475.
PR	23-SEP-1998;	98US-0101476.
PR	23-SEP-1998;	98US-0101477.
PR	23-SEP-1998;	98US-0101479.
PR	24-SEP-1998;	98US-0101738.
PR	24-SEP-1998;	98US-0101741.
PR	24-SEP-1998;	98US-0101743.
PR	24-SEP-1998;	98US-0101915.
PR	24-SEP-1998;	98US-0101916.
PR	29-SEP-1998;	98US-0102207.
PR	29-SEP-1998;	98US-0102240.
PR	29-SEP-1998;	98US-0102307.
PR	29-SEP-1998;	98US-0102330.
PR	29-SEP-1998;	98US-0102331.
PR	30-SEP-1998;	98US-0102484.
PR	30-SEP-1998;	98US-0102487.
PR	30-SEP-1998;	98US-0102570.
PR	30-SEP-1998;	98US-0102571.
PR	01-OCT-1998;	98US-0102684.
PR	01-OCT-1998;	98US-0102687.
PR	02-OCT-1998;	98US-0102965.
PR	06-OCT-1998;	98US-0103258.
PR	06-OCT-1998;	98US-0103449.
PR	07-OCT-1998;	98US-0103314.
PR	07-OCT-1998;	98US-0103315.
PR	07-OCT-1998;	98US-0103328.
PR	07-OCT-1998;	98US-0103395.
PR	07-OCT-1998;	98US-0103396.
PR	07-OCT-1998;	98US-0103401.
PR	08-OCT-1998;	98US-0103633.
PR	08-OCT-1998;	98US-0103678.
PR	08-OCT-1998;	98US-0103679.
PR	08-OCT-1998;	98US-0103711.
PR	14-OCT-1998;	98US-0104257.
PR	20-OCT-1998;	98US-0104987.
PR	20-OCT-1998;	98US-0105000.
PR	20-OCT-1998;	98US-0105002.
PR	21-OCT-1998;	98US-0105104.
PR	22-OCT-1998;	98US-0105169.
PR	22-OCT-1998;	98US-0105266.
PR	26-OCT-1998;	98US-0105693.
PR	26-OCT-1998;	98US-0105694.
PR	27-OCT-1998;	98US-0105807.
PR	27-OCT-1998;	98US-0105881.
PR	27-OCT-1998;	98US-0105882.

```

Query Match          68.8%; Score 5717.5; DB 21; Length 1523;
Best Local Similarity 66.6%; Pred. No. 1e-301;
Matches 1017; Conservative 217; Mismatches 277; Indels 15; Gaps 9;

Qy      3  GVGWQMLS-LSLGLVLA-ILNKVAPQACPAQCSGSGSTVDCHGLALRSVPRNPINRTERL 60
      |||  : : :| | | :|  |||  :| :| :| :| :| :| :| :| :| :| :|
Db      7  gvgaavrarlalalalasvlgppavacctkctcsaasvdchglglravrgiprnaerl 66

Qy      61 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRLNRNHLQLFLE 120
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      67 dldrnnitritkmdfaqlknrlvhlhdegngsvierqafgdiklerlrlnknklylpe 126

```

```
Qy      121 LLFLGTAKLYRLDLSNQIAIPKFAFRGAVDKNKLQDYNQISIEDGAFRALRDLEVL 180
       ||| | | | | | | | | | | | | | | : | | | | | | | | | | | | | | | | |
Db      127 llfgstpkltlrldlsenqigqiprkafrgaitdvknqlqdnnhisciedgafralrdleil 186

Qy      181 TLNNNNITRSLVSASFNHMPKLRTFRLHNSNNLYCDCHLAWLSDWLRRPRVGLYTCQMGPSS 240
       ||||| |:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      187 tlannnsirilvtsfnhmppkirtlrlshnhydcchlawsdlwrqrvtvgftlcmapv 246

Qy      241 HLRGHNVAEVQKREFCVCSDEEBHGQSFMAPSCSV--LHCPAACTCSNNIVDCRGKGLTEI 298
       |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      247 hlrgfnavdqvkkeyvcphaps-----eppscnansiscspctcsnnivdcrngklmei 301

Qy      299 PTNLPEITEIRLEQNTIKVIPPGAFSPYKKLRRIDLNNQISELAPDAFOGLRSLNSLV 358
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      302 panlpegiveirleqnikaipagafgtqykkikridiskngdislapdafggkslitslv 361

Qy      359 LYGNKITELPKSFEGLFSQLLLLLNANKINCLRVDAQDLHLNLLSLYDNKLQTIAGK 418
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      362 lygnkiteiakglfdglsiqllllnankinclrntfcdlglnllslsydnklqtiskg 421

Qy      419 TFSPLRAIQTMHLAQNPFICDCHLKWLADYLHTNPETSGARCTSPRRLANKRIGQISK 478
       | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      422 lfaplgisqtlhlagnpfvcdchlkwladyldnpdietsgarcssprrlankrisgisk 481

Qy      479 KFRCSGTEDYRSKLSGCDFADLACEKECRCEGTTVDCCSNQKNKIPEHIPOYTaelRLNN 538
       ||||| : |||| | | | | | | | | | | | | | | | | | | | | | |
Db      482 kfrccsgsedyrfsfecfmldvcpkekrcegtivdcscnqklvrishplepytvlrind 541

Qy      539 NEFTVLEATGIFKKLPQLRKINFNNKITDIIEGAFEGASGVNEILLTSNRLENVQHMF 598
       || : | | | | | | | | | | | | | | | : | | | | | | | | | | |
Db      542 nevsvleatgifkklpnlrkinnsnnikievregafdgaasvqmmltgnetvhgrvf 601

Qy      599 KGLESKLTLMLRSNRITCVGNDSEFIGLSSVRLLSYDNQITTVAPEGFDTHLSLTLL 658
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      602 rglsqglktlmrlsrnliscvsndtfaglssvrllsydnrittitpgafftlvslstinll 661

Qy      659 ANPFCNCYLAWLGWELRKRRIVTNGPCQKPFLKEIPIQDVAIQDFTCDDGNDNDSCLS 718
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      662 snpfcncchlawlgkwlrkrriavgpnrcqkpflkeipigdvaigtftc-dgneesscq 720

Qy      719 PLSRCTECTCLDTVVRCNSKGKLVLPKGIPTDVTLEYLDGNQFTLVPKELSNIYKHLLT 778
       ||| : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      721 lsprcceqtcmetvvrcnskglralrgmpkdvtleylegnhltavprelsalrhltl 780

Qy      779 DLSNNRISTLSNQSFNMQTLLTLILSYNRLRCIPPRTFDGLSKRLRLSHGNDISVVE 838
       |||| | | | : | | | | | | | | | | | | | | | | | | | | | |
Db      781 dlsnnsismlntytsfmsmshlstlilsynrlrcipvhafnglrsrvlithgndissve 840

Qy      839 GAFNDLSALSHLAIGANPLYCDCNMQWLSDWVKSEYKEPGIARCAGGEMADKLLLTPPS 898
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      841 gsfnditlshlagltnplhdcsrlwsewkagykegiarccssepmadrllltpt 900

Qy      899 KKFTCGQPVDVNILAKCNPCLSNPCKNDGTCSNDPVDFRYCTCPYGFGQCDDVIHACI 958
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      901 hrfcgckqpvdinivakcnacslsppckngtctqdgpvelracpysykgtkdtvpintci 960

Qy      959 SNPKCHKGGTCHLKEGEEDGFWCICADGFEGENEVENVDDCEDNDCENNSTCVDGINNYTC 1018
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      961 qnpcqhgtghlsdshdkdfscscplgfegqrceinpddcedndcennatcvdginnyc 1020

Qy      1019 LCPPEYTGELOCBEKLDFCAQDLNPCQHDSCILTPKGFKCDCTPGYVGHEHCDIDFDDCQD 1078
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1021 lcppnytgelcdevidhcpelnlcqheackipldkgfsccevpygsyklcetdnddcva 1080

Qy      1079 NKCKNGAHCTDAVNGYTCICEPYGSLFCEFSPPMVLPTSPCDFNDCQNAQQIVIRINE 1138
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1081 hkchrhaqcqvdingytctcpqgisfpgefhpmmvilqtspcdgyecqngaqcivrqqe 1140

Qy      1139 PICQCLPGYQGEXCEKLVSVNFINKESYLQIPSARKVRPQTNITLIQATDEDSGILLYKGD 1198
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1141 ptercpppgfajprceklietvnfygdksyvlasakvrpanisicvatdkngalllykgd 1200
```



```

RESULT      9
Y04137
ID   Y04137 standard; Protein; 1523 AA.
XX
AC   Y04137;
XX
DT   15-JUN-1999 (first entry)
XX
DE   Human slit 3 protein.
XX
KW   Human; slit-like protein; slit 3; slit 1; prevention; treatment;
KW   disease; spinal cord; thyroid gland; ovary; prostate; renal gland;
KW   small intestine; heart; trachea; thymus; lymph node;
KW   muscular system; colon.
XX
OS   Homo sapiens.
XX
FH   Key                Location/Qualifiers
FT   Peptide            1..17
FT                        /label= signal
FT   Protein            18..1523
FT                        /label= slit_3
XX
PN   JP11075846-A.
XX
PD   23-MAR-1999.
XX
PF   02-SEP-1997;      97JP-0236994.
XX
PR   02-SEP-1997;      97JP-0236994.
XX
PA   (ASAH ) ASahi Kasei Kogyo KK.
XX
DR   WPI; 1999-257695/22.
DR   N-PSDB; X19946.
XX
PT   New slit-like polypeptide - useful for prevention and treatment of
PT   diseases in spinal cord, thyroid gland, ovary, prostate, renal

```

```

Qy      3 GVGWQMLS-LSGLVLA-ILNKVAPQACPAQCSCSGSTVDCGHGLARLSVRPNRPRTERL 60
      ||| : : |:| || :|: ||| :|:| |::| ||| ||| |||
Db      7 gygaavrarlaralalasvlsppavacptkctcsaasvdcghglravprgiprnaerl 66

Qy     61 DLNGNNITRITKTDFAGLRHLRVQLMENKISTIERGAFQDLKELERLRLNRNHLQFPPE 120
      ||: ||||| ||||: ||| :|:| |||||: ||||| | ||: ||
Db     67 dldrnritritkmdfaglknlrvhlhednqvsviergafgdlkqlerlrlnknkqlvpe 126

Qy    121 LLLFLGTAKLYRLDLSENQIQAIKPRKAFRGAVDIKNLQLDYNOISCIEDGAFRALRDLEVL 180
      ||| | || ||||| ||||| |::| ||| | ||||| |||||:|
Db    127 llfqstpkltrldlseqigqgiprkafrgitdvknlqldnnhisciedgafralrdleil 186

Qy    181 TLNNNNITRLSVASFNHMPKLRTRFLHSNNLYCDCHLAWLSDWLKRKRPRVGLYTQCMGPS 240
      |||||:|: | |||||:| ||||:|||||||:| | :| | |
Db    187 tlnnnnisrlvtsfnhmpkirtlrhshlycdchlawlsdwlrqrrtvqgftlcmavp 246

Qy    241 HLRGHNVAEVQKREFVSCDEEBEGHSQFMAPSCSV---LHCPAACTCSNNIVDCRGKGLTEI 298
      |||| ||:|:|:|:| ||| : ||: ||||| ||||| ||
Db    247 hlrgfnvadvgkkeyvcaphs-----epsscnaansiscpsptcsnnivdcrgkglmei 301

Qy    299 PTNLPETITEIRLEQNTIKVIPPAGFSPYKKLRIDLSNNQISELAPDAFQGLRSLNSLV 358
      | ||| | |||||:| | ||: ||||| |||||: |||||:| |||
Db    302 panlpegiveirleqnsikaipagaftgykklkridsknqisdapdafgqlksltslv 361

Qy    359 LYGNKITELPKSLFEGLFSLQLLLLANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAGK 418
      |||||:| | ||:| ||||| ||||| ||||| ||||| |||||
Db    362 lygnkiteiakgfdglvslqlllnankinclrvtfgdlqnlnlslydnklqtiskg 421

Qy    419 TFSPLRAIQTMHLAQNPFCIDCHLKWLDYLHTNPIETSGARCTSPRLANKRIGQIKSK 476
      ||:|:|:|:|:| | | ||: ||||| ||||| ||||| |||||
Db    422 lfapqlsgitqlhlaqnpfvdcchkwladyldnpietsgarcssprlankrisqisk 481

Qy    479 KFRCSGTEDYRSKLSGDCFADLACPEKRCRCEGTVDCSNQKNLKIPEHIPQYTAELRLNN 538
      ||: |||||:| | || |||||:| |||||:| |||||:| |||||
Db    482 kfrcsgsedysrfsscfdmvlvcpekrcegtvdcsnqkvlripshlpeyvtdirlnd 541

Qy    539 NEFTVLEATGIFKKLPQLRKINFSNNKITDIEEGAFEGAGSVNEILLTSNRLENVQHMF 598
      || : ||||| |||||:| |||||:| |||||:| |||||:| |||||
Db    542 nevsvleatgiffklpnlrkinlnsnkikevregafdgaaqvqelmtgnqletvhgrvf 601

Qy    599 KGLESILKTLMLRSNRITCVGNSDFIGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLL 658
      :| : ||||| | || |||||:| |||||:| |||||:| |||||
Db    602 rglsqgkltlmlrsnlgcvsndtfaglsvrrlslydnrittpgaftlvslstinll 661

Qy    659 ANPFCNCNYLAWGEWLKKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGDDNDDSCS 718
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    662 snpfcncnchwlgkwlkrkrivsgnprcqkpfllkelpigdvaiqdfc-dgneesscq 720

Qy    719 PLSRCPTETCTLDTVVRCSNKGLKVLPGKIPRDVTELYLDGNQFTLVPKELSXYKHLTLI 778
      ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db    721 lsprcpegtcmetyvrcsnkgylralprgmkpdkvtelyegnhltavprelsalrhltil 780

Qy    779 DLSNNRISTLSNQSFNNMTQLLTLLSYNRLRCCIPPTFDGLKSLRLLSHGNDISVVPE 838
      |||||:| | |||||:| |||||:| |||||:| |||||:| |||||
Db    781 dlsnnsismtnvtfnsmshlsltlisvnrlicpvhafnlrslvltihandissvne 840

```

Qy 839 GAFNDLSALSHLAIGANPLYCDCNMQLSDWKSEYKEPGIARCAGPGEMADKLLLTPTS 898
Db 841 gsfnditslshlqgtphlhdcsrlwsewkygkpgiarccsspepmadrlllttpt 900

Qy 899 KKFTCQGPVDVNLAKCNPCSNPCNDGTCNSDPVDFYRCTCPYGFKGQDCDVPIHACI 958
Db 901 hrfgqkgpvdinivakcnacslspckngtctqdpvelryrcacpysykgkdcetvpintci 960

Qy 959 SNPCKHGGTCHLKEGEEDGFWICADGFECEGNEVNDVDCEDNDCENNSTCVDGINNYTC 1018
Db 961 qnpqhggtchlsdshkdgfcscplgfegqrceinpdcdndcennatcvgdginnyvc 1020

Qy 1019 LCPPEYTGEELCEKLDFAQDLNFCQHDSCILTPKGFKCDCTPGYGEHCDIDFDDCQD 1078
Db 1021 lcpnytgelcdevidhcvpelnlcqheakcpldkgfscecvpgysgkicetdndcva 1080

Qy 1079 NKCKNGAHTDAVNGYTCICPEGYSGLFCEFPMPVLPRTSPDCNFCQGAQCIVRINE 1138
Db 1081 hkrhgaqcvdtingytcctcpqgfsqfchppmvllqtspcdqyccqagacivvqge 1140

Qy 1139 PICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPNTITLQIATDEDSGILLYKGD 1198
Db 1141 ptcrcppgagprcekltitvfvkgdsyvelasakvrqanislqvatkdnlgillykd 1200

Qy 1199 KDHIARELYRGRVRSYDTSHPASAIYSVETINDGNFHVILLALDQSLSLSDVGGNPK 1258
Db 1201 ndplalelyqghvrlvydssppptvysvetvndggfshvelvtlntlnlvdkgtpk 1260

Qy 1259 IITNLSKQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHGCCIRNLYINSELQDFQK 1318
Db 1261 slgklqkqavnginsplylgiptstglsalrqgtdrplggfhgcihevinnelqdfka 1320

Qy 1319 VPMQT-GILPGCEPCHKVCAGHTCQPSQAGTCECQEGWMPCLDQRTNDPCLGNKCV 1377
Db 1321 lppqslgvsqgksc--tvckhglcrsvvekcrcrgwtgplcdqearpdlghrch 1378

Qy 1378 HGTCLPINAFSYSCCKLEGGHGVLCDEEEDLPNCPQAIKCKHGRSLGLQPYCECSSG 1437
Db 1379 hgkcvatgt-symkcaegygdldcnkndsanaacsfkchhgqchisdgqepcylcpg 1437

Qy 1438 YTGSDSCREISCRGRIYQKQGYAACQTTKKVSRLECRGGCAGGCGGGLRKRK 1497
Db 1438 fsghecgqenpclgqvrvirrrqkygascataaskvpimecrggc-gpccqptrskrrk 1496

Qy 1498 YSFECTDGSSFVDEVEKVKCGCTRC 1523
Db 1497 yvfqctdgssfveerhlecglac 1522

RESULT 10

Y27143

ID Y27143 standard; protein; 1496 AA.

XX

AC Y27143;

XX

DT 15-SEP-1999 (first entry)

XX

DE Human slit-3 mature protein (Seq ID No: 7 of JP11164690).

XX

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;
muscle; endocrine system.

XX

OS Homo sapiens.

XX

PN JP11164690-A.

XX

PD 22-JUN-1999.

XX

PF 05-DEC-1997; 97JP-0335435.

XX

PR 05-DEC-1997; 97JP-0335435.

XX

PA (ASAH) ASahi KASEI KOGYO KK.

XX

DR WPI; 1999-411830/35.

DR N-PSDB; X89163.

XX

PT New vertebrate slit protein - useful for diagnosis and treatment of
cancers in nerves, muscle and endocrine system

XX

PS Claim 6; Page 47-51; 102pp; Japanese.

XX

CC The invention relates to a vertebrate-derived protein containing an amino
acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein
has at least 55 % homology to one of sequences shown in Y27141-Y27143,
and has slit protein-like activity. The vertebrate slit proteins encoding
CC nucleic acid sequences have at least 60% homology to nucleic acid
sequences X89161-163. The vertebrate-derived proteins can be produced
recombinantly by transforming host cells with expression vectors
comprising the encoding nucleic acids. The proteins of the invention are
CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine
system.

XX

SQ Sequence 1496 AA;

Query Match 68.6%; Score 5702.5; DB 20; Length 1496;
Best Local Similarity 67.2%; Pred. No. 6.6e-301;
Matches 1008; Conservative 211; Mismatches 268; Indels 13; Gaps 7;

Qy 27 ACPAQCSGSGTVDCGLALRSVPRNIPRTERLDLGNNTITRTKTFAGLRHLRLVLQL 86
Db 6 acptkctcsaasvdchglrlavprgiprnerldldnnitritkmdfaglknlrvhl 65

Qy 87 MENKISTIERGAFQDLKELERLRLNRHLQFPPELLFLGTAKRLRLDSENQIQAIPRKA 146
Db 66 ednqvsviergafqdlkqlerlrlnknlgvlpellfqstpkrlrldsenqigqiprka 125

Qy 147 FRGAVIDKNLQLDYNQISCIEDGAFRALRDLEVLTLNANNITRLSVASFNHMPKLRTRFL 206
Db 126 frgtdvknlgldnnhisciedgafalrdleiltlannnislrvtsfnhmpkirtlrl 185

Qy 207 HSNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPHSLRGHNVAVQKRFVCSDEEGHQS 266
Db 186 hsnhlycdchlawlswlrqrtrvgftlmapvhlrgfnvadqkkeyvcphs---- 241

Qy 267 FMAPSCSV--LHCPAACTCSNNIVDCRGKGLTEIPTNLPEITIEIRLEQNTIKVIPPAGF 324
Db 242 -eppscnansiscspctcsnnivdcrgkglmeipalpegielrleqnsikaipagaf 300

Qy 325 SPYKRLRLDLSNQLSELAPDAFQGLRSLNLVLYGNKITELPKSLFEGFLSLQLLLN 384
Db 301 tgykklridlsknqisdiapafqglsltslvlygnkiteiakglfdglvslqllln 360

Qy 385 ANKINCLRVDAFQDLNLNLNLLSLYDNKLTIAKGTFSPLRAIQTMHQAQNPFCDCHLKW 444
Db 361 ankincrlrvntfdqlnlnllslydnkltiskglfaplsqitlhlagnpfvcdchlw 420

Qy 445 LADYLTNPITSGARCTSPRRANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADLACPE 504
Db 421 ladyldnpietsgarcssprlankrisqkskfrcsedyrsfscfmdlvce 480

Qy 505 KCRCEGTIVDCSNQKLNKIPHIQYTAELRLNNEFTVLEATGIFKLPQLRKINFSNN 564
Db 481 krccegtivdcsnqklvripshlpeyvtdrlndnevsvleatgikfklplnrkinfsnn 540

Qy 565 KITDIEGAFEGAGSVNEILLTNRLENVQHMKFGLESKLTMLRSNRITCVGNDSEFIG 624
Db 541 kikevregafdgasvqelmtgnqletvhrvfrglsglktmlrsnrlgcvsndtfag 600

Qy 625 LSSVRLSLYDNQITVAPGAFDYLHSLSTLNLANPNCNCLAWLGWLKRRKRVGN 684
Db 601 lssvrlslslydnrittitpgafitvslstlnlanspfnccnclawlgwlkrkrivsgn 660

Qy 685 PRCQKPYFLKEIPIDVAIQDFTCDGNDNDCSPLSRCTECTCLDTVVRCNSKGLKVL 744

Db 661 prcqkpfllkeipiqdvaigdfc-dgneesscqlsprceqctcmetvrcsnkgllral 719
Qy 745 PKGIPRDVTELYLDGNQFTLVPEKLSNYKHLTLIDLNNRISTLSNQSFNMTQLTLIL 804
Db 720 prgmprkdvteylegnhltavprelsalrhltidlsnnsismtntyfsnmshltlil 779
Qy 805 SYNRLRCIPPRFTDGLKSLRLSLHGNDISVVEGAFNDLSALSHLAIGANPLYCDNCMQ 864
Db 780 synrlrcipvhafnglrsrlvltlhgnidssvpegsfndltslshlaltgnplhcdcsir 839
Qy 865 WLSDWKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNILAKNPNCLSNPK 924
Db 840 wisewkagykepgiarccspsmadrilltptphrfqckpvdinivaknaclsspc 899
Qy 925 NDGTCNSDPVDFYRCTCPYGFQGDQCDVPHACISNPKHGGTCHLKEGEEDGFWICAD 984
Db 900 nngtctqdpvelrcacpysykgkdctvpintciqncpchggtchlsdshkgfscscl 959
985 GFEGENCEVNVDDCEDNDCENNSTCVGINNYTCLCPPEYTGELCEEKLDFAQDLNPO 1044
Db 960 gfegqrcelpddcedndcennatcvdginnyvcipppnytgelcdevdhcvelnclq 1019
Qy 1045 HDSKCLTPKFGKCDCTPGYVGEHCIDDFDCQDNKCKNGAHCDAVNGYTICPEGYSG 1104
Db 1020 heackpldkgfscecvpgysgkicetdndcvahkerhgaqcvdtingytcpcgfsq 1079
Qy 1105 LFCFSPMVLPRTPSPCDNFCQNGAQCIVRINEPICQCLPGYQGEKCEKLVSNFINKE 1164
Db 1080 pfcehppmvllgtqpcdqyccqngagcivvqgeptcrppgagprceklitnfvkgd 1139
Qy 1165 SYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDNDHIAVELYRGRVRSYDTGSHPSA 1224
Db 1140 syvelasakvrpaganislgvatdkdngilkygdndplalelyqghvrlvydssppt 1199
Qy 1225 IYSVETINDGNFHVILLALDQSLSLSDGNGPKIITNLSKQSTNFDSPLYVGGMPGKS 1284
Db 1200 vysvetvndgqfhsvelvtlnqtlnlvdkgtkslgklqkqpaqinsplylgiptst 1259
Qy 1285 NVASLRQAPQNGTSFHCIRNLYINSELQDFQKVPMT-GILPGCEPCHKVCAHGTQC 1343
Db 1260 glsalrqgtdrplggfhgcihevriinellqdkalppqslgvsppgksc--tvckhglcr 1317
Qy 1344 PSSQAGFTCECQEGWMPGLDQRTNDPCLGNKVHGTCLPINAFSYSCCKLEGHGVCLD 1403
Db 1318 svekdvsvcecrpgwtgplcdgeardpchlghhkgvatgt-symckcaegygdldc 1376
1404 EEEELFNPCQAIKCKHKGKRLSGLGQPYCECSSGYTGDSCDREISCRGERIRDYKQKG 1463
Db 1377 nkndsanaacsfkchhggchisdqgepyclcgpgfsgehcqgenpcqlqvrvirrkq 1436
Qy 1464 YAAQCTTKKVSRLCEGCGAGGCCPLRSKRRKYSPECTDGSFVDEVEKVVKGCTRC 1523
Db 1437 yaseataskvpimecrggc-gpgccqptrskrryvfqctdgssiveeverhlecyclac 1495

RESULT 11

Y04136

ID Y04136 standard; Protein; 1496 AA.

XX

AC Y04136;

XX

DT 15-JUN-1999 (first entry)

XX

DE Human slit 3 mature protein.

XX

KW Human; slit-like protein; slit 3; slit 1; prevention; treatment;
KW disease; spinal cord; thyroid gland; ovary; prostate; renal gland;
KW small intestine; heart; trachea; thymus; lymph node;
KW muscular system; colon.

XX

OS Homo sapiens.

XX

PN JP11075846-A.
XX
PD 23-MAR-1999.
XX
PF 02-SEP-1997; 97JP-0236994.
XX
PR 02-SEP-1997; 97JP-0236994.
XX
PA (ASAH) ASahi KASEI KOGYO KK.
XX
DR WPI; 1999-257695/22.
DR N-PSDB; X19946.
XX
PT New slit-like polypeptide - useful for prevention and treatment of
PT diseases in spinal cord, thyroid gland, ovary, prostate, renal
PT gland, small intestine, heart, trachea, thymus, lymph node, muscular
PT system and colon
XX
PS Claim 1; Page 17-21; 48pp; Japanese.
XX
CC The present sequence represents a human slit-like protein designated
CC slit 3. Slit-like proteins can be used for the prevention and the
CC treatment of diseases in spinal cord, thyroid gland, ovary, prostate,
CC renal gland, small intestine, heart, trachea, thymus, lymph node,
CC muscular system and colon.
XX
SQ Sequence 1496 AA;

Query Match 68.6%; Score 5702.5; DB 20; Length 1496;

Best Local Similarity 67.2%; Pred. No. 6.6e-301;

Matches 1008; Conservative 211; Mismatches 268; Indels 13; Gaps 7;

Qy 27 ACPAQSCSGSTVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTFAGLRHLRLVLQ 86
Db 6 acptkctcsaasvdchglrlavprgipraerldldrnnitritkmdfaglknlrvlhl 65
Qy 87 MENKISTIERGAFQDLKELERLNRNHLQLPELLFLGTAKLYRLDLSNQIAIPKA 146
Db 66 edngvsviergafgdlqlerlrlnknklqvlPELLFqstpktrldlsenqigipka 125
Qy 147 FRGAVIDKMLLDYNIQSCIEDGAFRALRDLEVLTLNANNITRISVASFNMHPKLRIFRL 206
Db 126 frgtdvknldnnhisciedgafalrdleiltlnnnnrisrlvtsfnhmpkirtlrl 185
Qy 207 HSNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPHSLRGHNAEVQKREFVCSDEEGHQS 266
Db 186 hsnhlycdchlawlswlrqrtrvgftlcmaphvlrgfnvadvgkkeyvcaphs---- 241
Qy 267 FMAPSCSV--LHCPAACTCSNNIVDCRGKGLTEIPTNLPEITTEIRLQNTIKVIPPAG 324
Db 242 -eppscnansiscspctcsnnivdcrgklmeipanalpegiveirleqnsikaipagaf 300
Qy 325 SPYKLLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLSLQLLLN 384
Db 301 tgykklridisknqisdapdafgqlksltslvlygnkiteiakglfdglvslqllln 360
Qy 385 ANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAGTFSPLRAIQTMHLAQNPFCIDCHLKW 444
Db 361 ankincrlrvntfqlqlnllslvdyndklqtliskglfaplsiqtlhlaqnpfvdchlk 420
Qy 445 LADYLHTNPIETSGARCTSPRLANKRIGQIKSKKFCSGTEDYRSKLSGDCFDLACPE 504
Db 421 ladyldnpietsgarcssprlankrisqskkfcsgsedysrfsfscmfdlvcp 480
Qy 505 KRCCEGTTVDCSNQKLNKIPHIQYTAELRLNNEFTVLEATGIFKLPQLRKINFSNN 564
Db 481 krccegtivdcnqklvripshlpeyvtldrlndevsvleatgikfklpnlrkinl 540
Qy 565 KITDIEGAFGAGSVNEILLTSNRLENVQHKMFGLSGLTLMLRSNRITCVGNDSPG 624
Db 541 kikevregaf3gaasvqelmgtgletvhgrvfrglsglktlmlrsnligcvsndtfag 600

Qy 625 LSSVRLSLYDNQITTVAPGAFDTLHLSLTLNLLNPFNCNCLAWLGEWLKRRIVTGN 684
Db 601 lssvrlslslydnrittttgafttlvslstlnllnspfnclawlgwlrkrivsgn 660

Qy 685 PRCQKPYFLKEIPIQDVAIQDFTCDGNDNSCSPSRCTECTCLDVTVRCSNKGKVL 744
Db 661 prcqkpyflkeipiqdvaiqdfctdgndnsccspsrcctectcldvtrcsnkgkvl 719

Qy 745 PKGIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQSFNMTQLTLIL 804
Db 720 pkgiprdvteylegndqftlvpkelsnykhltdlidsnnristlsnqsfntqltlil 779

Qy 805 SYNRLRCIPPRTFDGLSLRLSLHGNDSIVVPEGAFNDLSALSHLAIGANPLYDCNMQ 864
Db 780 synrlrcippftrfdglsrlslhgnsvvpegaftndlsalshlaiganplydcnmq 839

Qy 865 WLSDWKSEYKEPIARCAGPGEMADKLLLTTPSKKFTCQGPVDVNLAKNPNCLSNPCK 924
Db 840 wlsdwkseykepiarcagpgemadklllttppskkftcqpvdvnlaknnpclsnpck 899

Qy 925 NDGTCNSDPVDFYRCTCPYFGKQDCDVPFHACISNPKHGGTCHLKEGEEDGFWICAD 984
Db 900 ndgtcnsdpvdfyrctcpfygkqdcdvpfhacisnphggctchlkegeedgfwicad 959

Qy 985 GFEGENCEVNVDDCEDNDCENNCTVDGINNYTCLPPEYTGELCEELDFCAQDLNCPQ 1044
Db 960 gfegencevnnvddcedndcennctvdingnytclppeytgelceeldfcaqdlncpq 1019

Qy 1045 HDSKCLTPKGFKCDCTPGYVGEHCDIFDDQDNCKNGAHTDAVNGYTICPEGYSG 1104
Db 1020 hdsckcltpkgfkcdctpgyvgehcdifddqdnckngahtdavnngyticpegysg 1079

Qy 1105 LFCFSPMVLPRTPSPCDNFCQNGAQCIVRINEPICQLPGYQGEKCEKLVSVFINKE 1164
Db 1080 lfcfspmvlprtpspcdnfcqngaqcivrinepicqlpgyqgekekclvsvfinke 1139

Qy 1165 SYLQIPSAKVRPQTNTITLQIATDEDSGILLYKGDHIAVELYGRVRSYDTGSHPAS 1224
Db 1140 sylqipksavrpqntntitlqiattedsgillykgdhiavelygrvrsydtgshpas 1199

Qy 1225 IYSVETINDGNFHVILLALDQSLSLSDVGGNPKIITNLSKQSTLNFDSPLYVGGMPGKS 1284
Db 1200 ysvetindgnfnhfvillaldqslslsdvngnppkiitnlskqstlnfndsplyvngmpgks 1259

Qy 1285 NVASLRQAPGQNGTSFHGCIIRNLYINSELQDFQKVPMT-GILPGCEPCHKKVCAHGTQ 1343
Db 1260 nvaslrqapgqngtsfhgciirnllyinseleqdfqkvpmt-gilpgcepchkkvcahgtq 1317

Qy 1344 PSSQAGFTCECQEGWMLCDQRTNDPCLGNKCVHGTCLPINAFYSYCKLEGGVGLCD 1403
Db 1318 psqagftceceqegwmlcdqrdndpclgnkcvhgtclpinafysyckleggvglcd 1376

Qy 1404 EEEDLFNCPQAIKCKHGCRLSLGLQPYCECSSGYTGDSCDRSICRGERIRDYQKQOG 1463
Db 1377 nkdsanacsafkchhgcrlslglqpycecssgytgdscdrsisrgerirdyqkqog 1436

Qy 1464 YAAQCTTKVSRLECRGGCAGGQCCGLRKRKRSYFECTDGSFVDEVEKVVKCGCTRC 1523
Db 1437 yaaqcttkvsrlecrggcaggqccglrkrkrkyfeyctdgsfvdevekvkcgctrc 1495

RESULT 12

Y14142

ID Y14142 standard; Protein; 1523 AA.

XX

AC Y14142;

XX

DT 22-JUL-1999 (first entry)

XX

DE Human Slit protein sequence.

XX

KW Slit; human; diagnosis.

XX

OS Homo sapiens.

XX

PN WO9923219-A1.

XX

PD 14-MAY-1999.

XX

PF 29-OCT-1998; 98WO-US22845.

XX

PR 13-AUG-1998; 98US-0096420.

PR 31-OCT-1997; 97US-0063946.

XX

PA (OSIR-) OSIRIS THERAPEUTICS INC.

XX

PI Connolly T, Rajput B;

XX

DR WPI; 1999-337485/28.

DR N-PSDB; X61026.

XX

PT New human slit polypeptide and polynucleotide

XX

PS Claim 12; Fig 2; 90pp; English.

XX

CC This sequence is the human slit polypeptide of the invention.

CC The slit protein is useful for patients in need of the slit polypeptide,

CC and its antagonist is useful for patients in need of inhibition of

CC the slit polypeptide. Diagnosis of a disease or susceptibility to a

CC disease is achieved by determining the presence of a mutation in the slit

CC coding sequence, or determining the presence of slit by detecting

CC expression levels of the slit protein. Anti-slit antibodies are useful

CC for diagnosis of conditions associated with levels of slit protein.

XX

SQ Sequence 1523 AA;

Query Match 67.5%; Score 5611.5; DB 20; Length 1523;

Best Local Similarity 65.7%; Pred. No. 5.7e-296;

Matches 1002; Conservative 216; Mismatches 293; Indels 15; Gaps 9;

Qy 3 GVGQWMLS-LSLGLVLA-ILNKVAPQACQACSCSGSTVDCHGLALRSPVNPINRTERL 60

Db 7 gygaavrarlalalasvsgppavacptkctcsaasvdcgllgravrpgiprnaerl 66

Qy 61 DLNGNNITRIKTFDAGLRHLRLVLQNMENKISTIERGAFQDLKELERLRLNRHLQLFPE 120

Db 67 dldrnritritkmdfaglknlrvlhledngvsviergafqdlkqlerlrlnknklqvpe 126

Qy 121 LLFLGTAKLYRLDLSENQIQAIPKRAFAGVDIKNLQDYNQISIEDGAFRALRDLEVL 180

Db 127 llfqtspktrldlsenqigiprkafrgtdvknldnnhisciedgafralrdlell 186

Qy 181 TLNNNNITRLSVASFNHPKLRITRLHNSNLYCDCHLAWLSDWLRKRPRVGLYTCMGPS 240

Db 187 tlannnisrilvtsfnhmpkirtlrhshlycdchlawlswlrgtrvgftlmapv 246

Qy 241 HLRGHNVAEVQKREFVCSDEEFGHGFMAPSCSV--LHCPAECTSNNIVDCRGKGLTEI 298

Db 247 hlrgfnvadqkkeyvcapahs-----eppscnansiscspctcsnnivdcrgkglmei 301

Qy 299 PTNLPEITTEIRLEQNTIKVIPPAGFSPYKLLRIDLNNQISELAPDAFQGLRSLNSLV 358

Db 302 panlpegiveirleqnsikaipagafgykklridisknqisdiafpagqklsitlv 361

Qy 359 LYGNKITELPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLTQIAKG 418

Db 362 lygnkiteiakglfdglvslqlllnankinclrvtvtdlqnlslslydnkltqiskg 421

Qy 419 TFSPLRIQTMHLAQNPFICDCHLKWLDYLTNPDIETSGARCTSPRRLANKRIGQIKSK 478

Db 422 lfaplsiqtlhlaqnpfvcdchlkwladyldnpietsgarctssprlankrisqisk 481

Qy 479 KFRCSGTEDYRSKSLSGCFADLACPEKRCCEGTVDSCSNQKLNKIPHIQYTAELRLNN 538

Db 482 kfrcsgeydrsfscfmdlvcpckrccegtivdscsnqklvripshpeytdlrlnd 541

RESULT 13
W46966
ID W46966 standard; Protein: 1534 AA.

Query Match 67.3%; Score 5597; DB 19; Length 1534;
Best Local Similarity 65.3%; Pred. No. 3.5e-295;
Matches 992; Conservative 228; Mismatches 283; Indels 16; Gaps 5;

```

Qy 15 LVLAILNKVAPQACPAQSCSGSTVDCHGLALRSVPRNPRTNERLDDNGNNITRITKTD 74
   | : | : | | | | | : | : | : | | | | | : | : | : | | | | | | | |
Db 21 llwaaaawrlg|asacpalctctgttvdchgtglqalpknprnterlelngnnitrihknd 80

Qy 75 FAGRLHLRLVLQMLMENKISTIERGAFQDLKELERLRLNRNHLQFPBELLFGTAKLYRLDL 134
   | | | : | | | | | : | : | | | : | | | | | | | | | | | | | | | | | |
Db 81 faglkqlrlvlqlmenqigavergafddmkelerlrlnrnqhlmpellifqnnqalsrldl 140

Qy 135 SENQIQAIIPKRAFRGAVIDKINQLQDYNQISCIEDGAFRALRDLEVLTLNWNNTIRLSVAS 194
   | | | | | | | | | | | : | : | : | | | | | | | | | | | | | | | | | |
Db 141 senalqaiprkafrgatdklnlrldknqiscieegafalrglevltlnnnnittipvss 200

Qy 195 FNHMPKLRTFLHLSNNLYCDCHLAWLSDWLKRPRVGLYTCQMGPSHLRGHNVAEVQKRE 254
   | | | | | | | | | | | : | : | : | | | | | | | | | | | | | | | | | |
Db 201 fnhmpklrtflrhlshhlfcdchlawsgwlrqrptigtglqcsgpasrlgnlvaevqkse 260

Qy 255 FVCSDEEBEGHSFMAPSCSVL--HCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLE 312
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 261 fscsggqgeagr---vptctlssgscpamctcsnglvdcrgkgltaipanlpetmetirle 317

Qy 313 QNTIKVIPPGAFSPYKLLRRIDLNNQISELAPDAFOGLRSLNSLVLYGNKITELPKSLF 372
   | | | | | | | | | | | : | : | : | | | | | | | | | | | | | | | | | |
Db 318 lnlqksippqafspyrklrrldlssnclaelapdafoqlrslnslylvgnkitdlprafv 377

```

Qy 373 EGLFSLQLLLNANKINCLRVDAFDLHNLNLLSLYDNKLTQIAKGTFSPLRAIQTMHLA 432
Db 378 gglytlqllllnankincirpdafqdlqnlslslydnkqlslakgtftslraiqtlhla 437
Qy 433 QNPFICDCHLKLWADYLTNPETSGARCTSPRLANKRIGQIKSKKFRCS-----G 484
Db 438 qnpficdchlkwladflrtnpietsgarcasprlankrigqikskkfrcsakeqyfigp 497
Qy 485 TEDYRSKLSGDCPADLACPEKRCBEGTVDCSNQKLNKIPHIQYTAELRLNNEFTVL 544
Db 498 tedy--qlnsecnsdvcpkrcceanvvecsskltkiperipqstaerlnnneisil 555
Qy 545 EATGIFKKLPOLRKINFSNNKITDIEGAFEGASGVNEILLTSLNLENVQHRMFKGLESL 604
Db 556 eatgmfkkllthlkinlennkveledgafegaasvselhltanqlslrsgmfrgldgl 615
Qy 605 KTLMLRSNRITCVGNSDFIGLSSVRLSLYDNQITVAPGAFDTLHSLTLNLLANPFC 664
Db 616 rtlmlrnnrlscihndsfgrlnrvllslslydnqitvsgafdtlqslstlallanpfc 675
Qy 665 NCYLAWLGEWLKRRKIVTGNPRCQKPYFLKEIPQDVAIQDFTCDGNDNNSCPLSRCP 724
Db 676 ncqlawlggwlrrkivtgnprcqpdlrqlpqlqdvafdfrcceegqeggcprpccp 735
Qy 725 TECTCLDTVVRCNKGKLVLPKGIPTDTELYLDGNQFTLVPKELSNYKHLTLIDLSNR 784
Db 736 qecacldtvrcsnkhlralpkipnvtelyldgnqftlvpgqlstfkylqlvdlennk 795
Qy 785 ISTLSNQFSNMTQLLTLILSYNRLRCPRTFDGLSLRLSLHGNDISVPEGAFNDL 844
Db 796 isslsnssftnmsqltllslsynalqipplafggrslrlslhgndistlqegifadv 855
Qy 845 SALSHLAIGANPLYCDNCNQLSDWVKSEYKEPGIARCAGPGEMADKLLTTPSKFTCQ 904
Db 856 tslshlaiganplycdchrlwsswvktgykepgiarccagpgdmeqkllttpakkefcq 915
Qy 905 GPVDVNILAKNCPCLSNPKNDGTCSNDPVDYFRTCPYFGKQDCDVPFIHACISNPKCH 964
Db 916 gpptlavqakcdclsspcnqgtchndplevyrccacpsgygrdcvslnscssgpcen 975
Qy 965 GGTCHLKEGEDGFWCICADGFECEGENCEVNDCCEDNCCENNSTCVDGINNYTCLCPPEY 1024
Db 976 ggtchagagedapftscptgftgptcgvtndcdvhacanggvvcdgvnytcqclqy 1035
Qy 1025 TGLCEEEKLDFCAQDLNCPQHDSKILTPKFGKCDCTPGYVGEHCDIDFDDQCNCKNG 1084
Db 1036 egkaceqldlcsdpnlpcqhaeqvgtpdgrcecmppgyagdnscengdcdrhrcqng 1095
Qy 1085 AHCTDAVNGYTCICPEGYSGLFCEFSPPMVLPRISPCDNFDCNQAGCIVRINEPICQCL 1144
Db 1096 aqcmdevnsysclcaegysgqlceipphlpap-spcegtacqngancvddgnrpyvcql 1154
Qy 1145 PGYQGEKCEKLVSNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKHIAV 1204
Db 1155 pgfggpeceklsvnfvdrtlylqftdlqwnpranitlqvstaedngillyngdndhiav 1214
Qy 1205 ELYRGRVRSYDTGSHPASAIYSVETINDGNPHIVELLALDQSLSLSDGNGPKIITNLS 1264
Db 1215 elyqghrvrsydpqsyssalyaetindgqfhtvelvafdmvnlidqggsptmndfng 1274
Qy 1265 KQSTLNFDSPLYVGGMPGKSNVASLRQAPQNGTSFHKCIRNLYINSELQDFQKVPMTG 1324
Db 1275 khytlnseaplyvggmpvdvnsaafrlwqlngtgfhgicirnllylnnelqdfktqmkpg 1334
Qy 1325 ILPGCEPCHKKVCAHGTCPSSQAGFTCEQEGWMPGLCDQRTNDPCLGNKCVHGTCLPI 1384
Db 1335 vvpgepckrlyclhglcqpnatpgpmchceagvwlhcdqpadgpcghkcvhgycvpl 1394
Qy 1385 NAFYSYCKCLEGGHGLCDEEEDLPNCPQAIKCKHGLSLGLQPYCESSGYTGDSCD 1444
Db 1395 dalsyscqcqdgysgalcnqagalaepcrglclhghcqsagtkahcvcddpgfsgelce 1454
Qy 1445 REISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCGPLRSKRRKYSFECTD 1504

Db 1455 qesecrgdprdfhqvgrgyaiaqtrplswvecrgscpggqccggrlrrkrtfecsd 1514
Qy 1505 GSSFVDEVEKVKCGCTRC 1523
Db 1515 gtsfaeevekptkcgcalc 1533

RESULT 14

Y27144

ID Y27144 standard; protein; 1534 AA.

XX

AC Y27144;

XX

DT 15-SEP-1999 (first entry)

XX

DE Human slit-1 protein (Seq ID No: 11 of JP11164690).

XX

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve; muscle; endocrine system.

XX

OS Homo sapiens.

XX

PN JP11164690-A.

XX

PD 22-JUN-1999.

XX

PF 05-DEC-1997; 97JP-0335435.

XX

PR 05-DEC-1997; 97JP-0335435.

XX

PA (ASAH) ASahi KASEI KOGYO KK.

XX

DR WPI; 1999-411830/35.

XX

DR N-PSDB; X89161.

XX

PT New vertebrate slit protein - useful for diagnosis and treatment of

PT

cancers in nerves, muscle and endocrine system

XX

PS Disclosure; Page 51-57; 102pp; Japanese.

XX

CC The invention relates to a vertebrate-derived protein containing an amino acid sequence shown in Y27137 and Y27139. The vertebrate-derived protein has at least 55 % homology to one of sequences shown in Y27141-Y27143, CC and has slit protein-like activity. The vertebrate slit proteins encoding CC nucleic acid sequences have at least 60% homology to nucleic acid CC sequences X89161-163. The vertebrate-derived proteins can be produced CC recombinantly by transforming host cells with expression vectors CC comprising the encoding nucleic acids. The proteins of the invention are CC for diagnosing and treating cancer of the nerves, muscle and/or endocrine CC system.

XX

SQ Sequence 1534 AA;

XX

Query Match 67.2%; Score 5589; DB 20; Length 1534;
Best Local Similarity 65.2%; Pred. No. 9.5e-295;
Matches 991; Conservative 228; Mismatches 284; Indels 16; Gaps 5.

Qy 15 LVLAILNKVAPQACQACSCSGSTVDCHGLALRSVPRNIPRTERLDLNGNNITRITKTD 74

Db 21 llwaaawrlgasacpalctctgttdvchgthgqlprnterlelngnnitrlhkd 80

Qy 75 FAGLRHLRLVLQLMENKISTIERGAFQDLKELERLRLNRHLQLFPPELLFLGTAKLYRLDL 134

Db 81 faglkqlrvlqlmenqigavergafddmkelerlrlnrnqlhmlpellfngnntlrldl 140

Qy 135 SENQIQAIIPRAFRGAVDIKNLQDLYNQISCIEDGAFRALRLDLEVLTLNNNNITRLSVAS 194

Db 141 senaigalprkafrgatdlnlrldknqiscieegafalrglevltlnnnnitrtpvss 200

Qy 195 FNHMPKLRTFKLHNSNLYCDCHLAWLSDWLRRPRVGLYTQCMGPHLRGNVAEVQKRE 254

Db 195 FNHMPKLRTFKLHNSNLYCDCHLAWLSDWLRRPRVGLYTQCMGPHLRGNVAEVQKRE 254

Query Match : 67.2%; Score 5589; DB 20; Length 1534;
Best Local Similarity 65.2%; Pred. No. 9.5e-295;
Matches 991; Conservative 228; Mismatches 284; Indels 16; Gaps 5;

```
Qy      15 LVLAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRTERLDLNGNNITRIKTD 74  
       | : | : | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db     21 llwaawarlgasacpalctctgttvdchgtglgaipkniprnterlelngnnitrihknd 80  
  
Qy     75 FAGLRHLRLVLQLMENKISTIERGAFODLKELERLRLNRNLQLPPELLFLGTAKLYRLDL 134  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | |  
Db    81 faglkqlrvlqlmenqigavergafddmkelerlrlnrnqhlmpellfqnnqalsridl 140  
  
Qy    135 SENQIQAIPIPKRAFGRVAIDIKNLQLDYNIQISCIEDAGFARLRDLEVLTLNANNITRLSVAS 194  
       | | | | | | | | | | : | | | : | | | | : | | | | : | | | | : | : |  
Db   141 senaiqaiprkrafrgatdknlrlldkngiscieegafalrglevltlannnittipvss 200  
  
Qy    195 FNHMPKLRTFRHLSNNLYDCDHLAWLSDWLKRPRVGLYTQCMGPSHLRGHNVAEVQKRE 254  
       | | | | | | | | | | : | | | | | | | | | | : | | | | : | | | | |  
Db   201 fnhmpklrtfrlhslnhlfcdchlawsqwlrqpteltfqtcsgpsalrglnvaevqkse 260  
  
Qy    255 FVCSDDEEHHQSFMAPSCSVLT--HCPPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLE 312  
       | | | : | : | : | : | : | | | | | | | | | | | | | | | | | | | |  
Db   261 fscsggqeagr---yptctlssggscpmactcsngivdcrgkgлтаipanlpetmteirle 317  
  
Qy    313 QNTIKVIPPGAFSPYTKLRRIDLNNQISELPADFAOGLRSLNSLVLYGNKITELPKSLF 372  
       | | | | | | | | | | : | | | | | | | | | | : | | | | | | | | | | : |  
Db   318 lngiksippgafspyrklrridlsnnqiaelapdafgqlrslnslvlygnkitdprgvf 377  
  
Qy    373 EGLFSQLLLLLNANKINCLRVDAFDLHNLNLLSYDNKILQTIAGFTFSPLRAIQTMLHA 432  
       | | : | | | | | | | | : | | | | | | | | : | : | | | : | | | | : | | |  
Db   378 ggyltqllllnankincirpdfadqlnslslsydnkiqslakgtftsiraigtqhla 437  
  
Qy    433 QNPFI CDCHLKWLDYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFCRS-----G 484  
       | | | | | | | | | | : | | | | | | | | : | | | | | | | | : | | | | |  
Db   438 qnpficdclnkwladflrtntpietsgarcasprrlankrigqlskkkfcrcsaqeyfigp 497  
  
Qy    485 TEDYRSKLSGDCFADLACEKRCCEGTTVDCSNQLNKIPEHIPQYTAELRLNNEFTVL 544  
       | | | | : | : | : | : | : | | | | | | : | | | | | | | | : | | | | |  
Db   498 tedy--qlnsecsdvvcpkhkrceanvvecssklktkiperipqstaerlrnnneisil 555  
  
Qy    545 EATGIFKKLPQLRKINFSSNNKIIDIEGAPEGASGVNEILLTSNRLENVQHMKFKGLESL 604  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
Db   556 eatgmfkklthlkinlsnnkvseiedafegaasvselhltanqlsirsgmfrgldgl 615  
  
Qy    605 KTLMLRSNRITCVGNDISFGLSSVRLLSLEDYNQITTVPAGAFDTLHSLSTLNLANPFNC 664  
       | : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | |  
Db   616 rtlmlrnnriscihndsftglrnvrllslydnqittvspgafdtqlslstlnlanpfnc 675  
  
Qy    665 NCYLAWLGEWLRKKRIVTGNPCQKPFLKEIPIQDVAIQDFTCDDGNDNNSCPLSRCP 724  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
Db   676 ncqlawlggwlrkrkivtgnpccnqpdfirqlpqdvafpdrfceegggeggclprpqcp 735  
  
Qy    725 TECTCLDTVVCRSNGKGLVLPKGIPRDVTELYLDGNQFTLVPKELSINYKHLTLDLSNNR 784  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
Db   736 gecacldtvvcrsnkhrlalpkqipnvtyeldgnqftlvpgqlstkfylqlvdlssnk 795  
  
Qy    785 ISTLSNQSFSSNMTOQLTLILSYNRLCIPPPTFDGLKSLRLLSHGNDISVVEGAFNDL 844  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
Db   796 isslsnssftnmqlttlilsynalcqipplafqglrsirlslshgndistlqegifadv 855  
  
Qy    845 SALSHLAIGANPLYCDNCMNQWLSDWVKSEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQ 904  
       | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |  
Db   856 tslshlaiganplycdchlrlwssvwtktykepggiarcaggpdmegklllttpakkfecq 915  
  
Qy    905 GPVDVNILAKCNPLSNPCKNDGTCNSDPVDFYRCTCPYGFGQDCDVPIHACISNPCKH 964  
       | | : | : | | | : | | | : | | | : | | | : | | | : | | | : | | | |  
Db   916 gpptlavqakcdclsspccnggtchnplevyrcacpsgykgdrcevsinscsgpcen 975  
  
Qy    965 GGTCHLKEGEEDGFWCICADGFEGENCEVNVDDCEDNDCENNSTCVGDINNYYTCLCPPEY 1024  
       | | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | | |  
Db   976 ggtchaqgedapftcscptgfegptcgvtddcvdhacanggvvcvgnytcqcplay 1035  
  
Qy    1025 TGELCEEKLDFOCAQDLNPCQHSKCILTPKGFGKCDCTPGYVGHECHDIFDDCQDNKCKNG 1084  
       | | : | : | | | : | | | : | | | : | | | : | | | : | | | : | | | |  
Db   1036 egkaceqvlidlcspdlncpcheaqcvtpdpgrcrecmqpyagdncsenaddcdrhrcaq 1095
```

```

Qy 1085 ABCTDAVNGYTCICPEGYSGLFCFSPPMVLPRTPSPCNDFDCQNGAQCIVRINEPICQCL 1144
| | | | | : | : | | | | | | | | | | : | : | : | | | | | : | : | | |
Db 1096 aqcmdevnsysclcaegysgqlceipphlpak-spcegtceqngancvdqgnrpvcqcl 1154

Qy 1145 PGYQGEKCEKLSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKHIAV 1204
| | : | : | | : | | | : : : : | | | : | : | | | : | | | | | | | | | |
Db 1155 pfgfpgpeceklslsvnfvdrtqylqftldqgnpranitlqvstaedngillyngdndhiav 1214

Qy 1205 ELYRGRVRSYDTGSHSPAIYSVETINDGNFHIVELLALDQSLSLSDVGNGPKIITNLS 1264
| | : | | | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 1215 elyqghvrvsydpqsgypssalsyaetindgqgfhtvelvafdgqmvnlsidggsptmdnfg 1274

Qy 1265 KQSTLNFSDPLYVGGMPGKSNVASLRQAPQNGTSFHHGCI RNLYINSELQDFQKVPMTG 1324
| | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1275 khytlneaplyvggmpvdrvnsaafrlwqlngtghfghcicrnlylannelqdfktqmkpg 1334

Qy 1325 ILPGCEPCHKKVCAHGTCQPSQAGFTCECQEGWMGPLCDQRTNDPLCNKGVHGTCLPI 1384
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1335 vvpqcepcrklychlghcqpnatpgpmchceagvvlhcdqpadqpcghghckvhgqcvpl 1394

Qy 1385 NAFYSCKCLEHGGVLCDEBEDLFPNCQAIKCHKGKRLSGLGOPYCECSSGYTGDSCD 1444
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1395 dalsysccqcdgysgalcnagalaepcrglqlghhcgasgtgkahcvcdqpfsgelce 1454

Qy 1445 REISCRGERINDYYQKQGYAACQTTKKVSRLECRGGCAGGCCGPLSRKRKYSPECTD 1504
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1455 qesecrgdpvrdfhvgvrgyalcqttrplswwecrgscpgggccqglrlkrkftfecsd 1514

Qy 1505 GSSFVDEVEKVVKCGCTRC 1523
| : | | | | | | | | | |
Db 1515 gtsfaeevekptkcgcalc 1533

```

Search completed: January 22, 2001, 12:13:45
Job time: 1382 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 11:52:03 ; Search time 325.28 Seconds
(without alignments)
318.337 Million cell updates/sec

Title: US-09-540-245A-2
Perfect score: 8316
Sequence: 1 MRGVGNQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Sequences: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0;
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5728	68.9	1523	2	T13953
2	5578.5	67.1	1531	2	T42218
3	5530	66.5	1025	2	T42626
4	3486	41.9	1480	2	A36665
5	3475.5	41.8	1469	2	B36665
6	1318	15.8	333	2	T34555
7	1115.5	13.4	530	2	A31640
8	1006	12.1	601	2	T22025
9	818	9.8	2703	1	A24420
10	768	9.2	1220	2	A56136
11	759	9.1	2524	2	A35844
12	757	9.1	2352	2	T30201
13	747.5	9.0	2531	2	S18188
14	745	9.0	2531	2	A46019
15	736	8.9	1203	2	A49175
16	732.5	8.8	2437	2	S42612
17	732	8.8	2471	2	A49128
18	731.5	8.8	2555	2	A40043
19	731	8.8	1064	2	A40136
20	731	8.8	2531	2	T31070
21	722	8.7	2318	2	S45306
22	719.5	8.7	2321	2	S78549
23	691.5	8.3	1964	2	T09059
24	660	7.9	861	2	A48825
25	634.5	7.6	2139	2	A35672
26	625.5	7.5	1404	2	A36666
27	625.5	7.5	1408	2	S16148
28	606	7.3	1429	2	S06434
29	592	7.1	1389	2	T13852

30	586	7.0	1385	2	T13887	tlr protein - frui
31	568.5	6.8	728	2	I50719	C-Delta-1 - chicke
32	549.5	6.6	570	2	A48836	fibropellin C prec
33	543	6.5	1295	2	A32901	glpl protein precu
34	535.5	6.4	473	2	A56175	adhesive plaque pr
35	523	6.3	1372	2	T25933	hypothetical prote
36	518.5	6.2	722	2	I48324	DELTA-like 1 - mou
37	493.5	5.9	603	2	JC6128	insulin-like growt
38	491	5.9	833	2	S19087	gene Delta protein
39	490	5.9	1091	2	A58532	glial cell membran
40	489	5.9	603	2	JC1282	insulin-like growt
41	484.5	5.8	605	2	A41915	insulin-like growt
42	480	5.8	832	2	A31246	neurogenic protein
43	479.5	5.8	605	2	JC5239	insulin-like growt
44	478.5	5.8	880	2	S00670	neurogenic repetit
45	478	5.7	3034	2	T14119	seven-pass transme

ALIGNMENTS

RESULT 1
T13953
MEGF5 protein - rat
N;Alternate names: slit protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T13953
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like moti
A;Reference number: Z14126; MUID:98360089
A;Accession: T13953
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1523 <NAK>
A;Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C;Genetics:
A;Gene: MEGF5

Query Match 68.9%; Score 5728; DB 2; Length 1523;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1015; Conservative 223; Mismatches 265; Indels 14; Gaps 9;

Qy	11	LSLGLVLA-ILNKVAPQACPAQCSGSDVCHGLALRSVPRNIPRTERLDLNGNNITR	69
Db	16	LALALALASILSGPPAAACPTKCTCSAASVCHGLGLRAVPRGIPRNERLDLDRNNITR	75
Qy	70	ITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRNLRNLQFLPELFLGTAKL	129
Db	76	ITKMDFTGLKRLVHLLEDNQVSVIERGAFQDLKQLERLRNLRNLQFLPELFLQSTPKL	135
Qy	130	YRLDLSENQIQAIIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITR	189
Db	136	TRLDLSENQIQGIPKAFRGVTGVKNLQLDNNHISCIEDGAFRALRDLEILTLNNNNISR	195
Qy	190	LSVASFNHMPKLRTRFLSNLNYCDCHLAWLSDWLKRPRVGLYTQCMGPHSLRGHNVAE	249
Db	196	ILVTSFNHMPKIRTLRLSNLNYCDCHLAWLSDWLQRRTIGQFTLCMAVHLRGFSVAD	255
Qy	250	VQKREFVCSDEEGHQSFMAPSCSV--LHCPAACTCSNNIVDCRGKGLTEIPTNLPETIT	307
Db	256	VQKKEYVC---PGPHS-EAPACNANSLSCPSACSCSNNIVDCRGKGLTEIPANLPEGIV	310
Qy	308	EIRLEQNTIKVIPPAGAFSPYKLRRLIDLSNNQISELAPDAFQGLRSLNSLVYGNKITEL	367
Db	311	EIRLEQNSIKSIPAGAFIQYKLRRLIDLSNNQISELAPDAFQGLRSLNSLVYGNKITEL	370
Qy	368	PKSLFEGFLFSQQLLLNANKINCLRVDAFDLHNLNLSLYDNKLQTIAGKFTSPLRAIQ	427
Db	371	PKGLFDGLVLSQQLLLNANKINCLRVNTFDLHNLNLSLYDNKLQTIAGKFTSPLRAIQ	430

Qy 428 TMHLAQNPFICDCHLKWLDYLTNPITSGARCTSPRLANKRIGQIKSKKFRCSGTED 487
Db 431 TLHLAQNPFVCDCHLKWLDYLTNPITSGARCTSPRLANKRISQIKSKKFRCSGTED 490

Qy 488 YRSKLSGDCFADLACPEKCRCEGTTVDSCNQKLNKIEHIPQYTAELRLNNEFTVLEAT 547
Db 491 YRNFSSECFMDLVCEKCRCEGTTVDSCNQKLSRIPSHLPEYTTDLRLNDNDIAVLEAT 550

Qy 548 GIFKKLPQLRKINFSNNKITDIEEAGFEGASGVNEILLTSNRLNVQHKMFGLKSLKTL 607
Db 551 GIFKKLPNLRKINLSNNRIKEVREGAFDGAAGVQELMTGNQLETHGRMFRGLSLKTL 610

Qy 608 MLRSNRITCVGNDISGLSSVRLSLYDNIQTTVAPGADTLHSLSTLNLLANFPNCNCY 667
Db 611 MLRSNLSICVNDITAGLSSVRLSLYDNRITISPGATTLSLSTLNLLANFPNCNCY 670

Qy 668 LAWLGEWLKRRIVGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNDSCLSPRCPTC 727
Db 671 MAWLGEWLKRRIVGNPRCQKPYFLKEIPIQDVAIQDFTC-EGNEENSQSLSPRCPEQC 729

Qy 728 TCLDTVVRCSNKLGLVLPKGIPIRDTVLYLDNQFTLVPKLSNYKHLTLIDLNNRIST 787
Db 730 TCVETVVRCSNRLGLTLPGKMPKDVTELYLEGHNLTAVKELSTPRQLTLIDLNNISISM 789

Qy 788 LSNQSFNMTQLTILSYNRLRCIPPTFDGLSKRLSLHGNDSIVVPEGAFNDLSAL 847
Db 790 LTNHTFSNMHSLTILSYNRLRCIPVHAFNGLSLAVLTSLHGNDSIVVPEGSFNDLSAL 849

Qy 848 SHLAIGANPLYCDCHLWLDVWVSEYKEPGIARCAAGPEMADKLLTTPSKKFTQCGPV 907
Db 850 SHLAGINPLHCDCHLWLDSEWIKAGYKEPGIARCSSPESMADRLTTPTHRFQCKGPV 909

Qy 908 DVNILAKNPLSNKNDGTCNSDPVDFYRCTCPYFGKQDQVPIHACISNPKKHGT 967
Db 910 DINIVAKNACLSSCKNNGTCSQDPVQYRCTCPYSYKGDCTVPINTCVQNPQKHGT 969

Qy 968 CHLKEGEEDGFWICADGFEENCEVNVDDCEDNENNSTCVDGINNYTCLPPEYTG 1027
Db 970 CHLSHSHRDGFSCLGFEQRCINPDCCEDNDCENSATCVDGINNYACVCPNYTG 1029

Qy 1028 LCEEKLDFAQDLNQCQHSKCLITPKGFKDCTPGVYGEHCIDFDQDCKNKGARH 1087
Db 1030 LCDEVIDYCPENNTQHEAKISLKGFRCEVPGYSKGLCTDNDQVAKHGRHQAQ 1089

Qy 1088 TDVANGYTCICPEGYGLFCEFSPPMVLPTSPDNCQNGAQCIVRINEPICQLPGY 1147
Db 1090 VDAVNGYTCICQGSGLFCEHPPMVLQTSPTDQYQNGAQCIVVQEPCTCRCPGF 1149

Qy 1148 QGEKCEKLVSVNFINKESYLQIPSAKVRPQNTILQIATDEDSGILLYKGDHIAVELY 1207
Db 1150 AGPRCEKLTIVNFVKDSTVELASAKVRPQANISLQVATDKDNGILLYKGDNDPLALELY 1209

Qy 1208 RGRVRSYDTGSHPSAIYSVETINDGNFHVALLDQSLSLSDGNGNKIITNLSKQS 1267
Db 1210 QSHVRLVYDLSLSPPTTVYSVETVNDGQFHSVELVMLNTLNLVVDKGAPSLGKLQKQP 1269

Qy 1268 TLNFDSPLYVGMGPKSNVASLRQAPGQNGTSFHCIRNLYINSELQDFQKVPMT-GIL 1326
Db 1270 AVGINSPLYLGGIPTSTGLSALRQAGADPLGFGHCHEVRINNELQDFKALPQSLGVS 1329

Qy 1327 PGCEPHKKVCAHGTCQPSQAGFTCEQEGWGPCLDQRTNDPCLGNKCVHGTCLPINA 1386
Db 1330 PGCKSC--TVCRHGLCRSEKDSVVCHEPGWTGLDQEAQDPLGHSCHSGTCV-ATG 1386

Qy 1387 FSYSCKLEHGGVLCDEEDLNPQAIKCKHKGKRLSGLPYCESSGYTGDSCDRE 1446
Db 1387 NSYVCKCABGYEGPLCDQKNSANACSAFKHHGQCHISDRGEYCLCPGFSGNHCEQE 1446

Qy 1447 ISCRGERIRDYQKQGYAACQTTKVRSLRECRGCGAGGCCPLRSKRRKYSPECTDGS 1506
Db 1447 NPCLGEIVREAIRRQKDYASCATASKVPIMVCRGCG-GSQCCQPIRSKRRKIVFQCTDGS 1505

Qy 1507 SFVDEVEKVKCGCTRC 1523

Db 1506 SFVEEVERHLBCGCREC 1522

RESULT 2
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF4

Query Match 67.1%; Score 5578.5; DB 2; Length 1531;
Best Local Similarity 65.0%; Pred. No. 0;
Matches 988; Conservative 230; Mismatches 282; Indels 19; Gaps 6;

Qy 15 LVLAILNKVAPQACPAQCSGSDVDCCHGLALRSVPRNIPRTERLDLNGNNITRIKTD 74
Db 21 LLWAAAWRLGATACALCTCTGTVDCHGTGLQAIKPNIPRTERLELNGNNITRIKTD 80

Qy 75 FAGRLHRLVLQLMENKISTIERGAFQDLKELERLRLNRNLQLFPPELLFLGTAKLYRLDL 134
Db 81 FAGLKQLRLVLQLMENQIGAVERGAFFDMKELERLRLNRNLQLFPPELLFQNNALSRLDL 140

Qy 135 SENQIAIPKAFRGAVIDIKNLQDYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVAS 194
Db 141 SENSLQAVPRKAFRGAIDLKNLQDYNQISCIIEGAFRALRGLEVLTLNNNNITIPVSS 200

Qy 195 FNMHPKRLTFRHLSNLYCDCHLAWLSDWLRKRPRVGLYTCQMGPSHLRGNVAEVQKRE 254
Db 201 FNMHPKRLTFRHLSNLYCDCHLAWLSDWLRKRPRVGLYTCQMGPSHLRGNVAEVQKSE 260

Qy 255 FVCSDEEHHGQSFMAPSCSVL--HCPAACTSNINIVDCRGKGLTEIPTNLPTETIEIRLE 312
Db 261 FVCSGGGEAAQ---VPACTSSGSCPAMCSNGIVDCRGKGLTAIPANLPTETIEIRLE 317

Qy 313 QNTIKVIPPAGAFSPYKLRRLIDLSNNQISELAPDAFGQLRSLNLYGNKITELPKSLF 372
Db 318 LNGKISPPAGAFSPYKLRRLIDLSNNQIAETAPDAFGQLRSLNLYGNKITELPRGVF 377

Qy 373 EGLFSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLQTIAGKTFSPRLAIQTMLHA 432
Db 378 GGLYTLQLLLNANKINCLRVDAFQDLNLSLYDNKLQTIAGKTFSPRLAIQTMLHA 437

Qy 433 QNPFICDCHLKWLDYLTNPITSGARCTSPRLANKRIGQIKSKKFRCS-----G 484
Db 438 QNPFICDCHLKWLDYLTNPITSGARCTSPRLANKRIGQIKSKKFRCSAKEQYFIPG 497

Qy 485 TEDYRSKLSGDCFADLACPEKCRCEGTTVDSCNQKLNKIEHIPQYTAELRLNNEFTVL 544
Db 498 TEDYH--LNSECTSDVACPHKCRCEASVVCESGLKLSKIPERIPQSTELRLNNEFTVL 555

Qy 545 EATGIFKKLPQLRKINFSNNKITDIEEAGFEGASGVNEILLTSNRLNVQHKMFGLKSL 604
Db 556 EATGLFKKLSHLKINLSNNKVSEIDGTPEGATSVSEHLTANQLSVRSQMFRLDGL 615

Qy 605 KTLMLRSNRITCVGNDISGLSSVRLSLYDNIQTTVAPGADTLHSLSTLNLLANFPNC 664
Db 616 RTMLRLNRNISCINHSFTGLRNVRLSLYDNIHITISPGADTLQALSTLNLLANFPNC 675

[illegible]

RESULT 3
T42626

secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N:Alternate names: neurogenic extracellular slit protein
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T42626
R:Holmes, G.P.; Negus, K.; Burrridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A:Title: Distinct but overlapping expression patterns of two vertebrate slit homologs in
A:Reference number: Z22177; MUID:99279238
A:Accession: T42626

A;Status: preliminary; translated from GB/EMBL/DBBJ
A;Molecule type: mRNA
A;Residues: 1-1025 <HOL>
A;Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C;Genetics:
A;Gene: Slit2

Query Match 66.5%; Score 5530; DB 2; Length 1025;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 983; Conservative 25; Mismatches 17; Indels 0; Gaps 0;

Qy	501	ACPEKCRCEGTTVDCSNQRLNKIPEHIPQYTAELRLNNNEFTVLEATGIPFKLPQLRKIN	560
Db	1	ACPEKCRCEGTTVDCSNQRLNKIPDHIPQYTAELRLNNNEFTVLEATGIPFKLPQLRXIN	60
Qy	561	FSNNKITDIEEGAFEGASGVNEILLTNSRLENVQHKMFKGLESKLTMLRSNRITCVGND	620
Db	61	FSNNKITDIEEGAFEGASGVNEILLTNSRLENVQHKMFKGLESKLTMLRSNRISCVGND	120
Qy	621	SFIGLSSVRLLSLYDNQITTVAPGAFDTLHSLSTLNLLANPFCNCYLAWLGEWLKKRI	680
Db	121	SFIGLGSVRLLSLYDNQITTVAPGAFDXLHSLSTLNLLANPFCNCHLAWLGEWLRRKI	180
Qy	681	VTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNNSCPSLRCPTECTCLDTVVRCSNKG	740
Db	181	VTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNNSCPSLRCPSECTCLDTVVRCSNKG	240
Qy	741	LKVLPGKIPRDVTLEYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQFSNMTQLL	800
Db	241	LKVLPGKIPKDVTELEYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQFSNMTQLL	300
Qy	801	TLILSYNRLRCIPPRTFDGLKSLRLLSHGNDISVVPPEGAFNDLSALSHLAIGANPLYCD	860
Db	301	TLILSYNRLRCIPPRTFDGLKSLRLLSHGNDISVVPPEGAFNDLSALSHLAIGANPLYCD	360
Qy	861	CNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLTTPSKKFTCQGPVDVNI LAKCNPCLS	920
Db	361	CNMQWLSDWVKSEYKEPGIARCAGPGEMADKLLTTPSKKFTCQGPM DITIQACNPNCLS	420
Qy	921	NPKKNDGTCSNDPVDYFRTCTCPYFGKQDCDVIHACISNPKKHGGTCHLKEGEDGFWC	980
Db	421	NPKKNDGTCSNDPVDYFRTCTCPYFGKQDCDVIHACISNPKKHGGTCHLKEGENAGFWC	480
Qy	981	ICADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFAQDL	1040
Db	481	TCADGFEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFAQDL	540
Qy	1041	NPCQHSKCILTPKGFKCDCTPGYVGEHCDIFDDCQDNKKNKAHCTDAVNGYTCICPE	1100
Db	541	NPCQHSKCILTPKGFKCDCTPGYVGEHCDIFDDCQDNKKNKAHCTDAVNGYTCVCEP	600
Qy	1101	GYSLGFCEFSPPMVLPRTPSPCNDNFCQNGAQCIIRINEPICQLPGYQGEKCEKLSVNF	1160
Db	601	GYSLGFCEFSPPMVLPRTPSPCNDNFCQNGAQCIIRINEPICQLPGYLGEKCEKLSVNF	660
Qy	1161	INKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHI AVELYRGRVRSYDTGSH	1220
Db	661	VNKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHI AVELYRGRVRSYDTGSH	720
Qy	1221	PASAIYSVETINDGNFHI VELLALDQSLSLSDVGGNPKIITNLSKQSTLNFDSPLYVGGM	1280
Db	721	PASAIYSVETINDGNFHI VELLTLDSSLSLSDVGGSPKVIITNLSKQSTLNFDSPLYVGGM	780
Qy	1281	PGKSNVASLRJAPGQNGTSFHGCI RNLYINSELQDFQKVPMTGILPGCEPCHKKVCAHG	1340
Db	781	PGKNNVASLRJAPGQNGTSFHGCI RNLYINSELQDFRKMPMQTLGILPGCEPCHKKVCAHG	840
Qy	1341	TCQPSSQAGFTCECGEGWGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCCKLEHGHSV	1400
Db	841	MCQPSSQSGFTCECGEGWGPLCDQRTNDPCLGNKCVHGTCLPINAFSYSCCKLEHGHSV	900
Qy	1401	LCDEEEDLNFNCOAIKCHKGKCRSLGOLPGYCEC SSGYTGDSCDREISCRGERIRDYYOK	1460

Db 1156 FVHNNSFVELEPLRTRPEANVTIVFSSAEQNGILMYDGQDAHLAVELFNGRIRVSYDVG 121

Oy 985 GFEQENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLDFAQDLNPC 1043
| : | | |||| : ||:||:||: |: | : | : ||
Dh 976 CYGCACDCTENDDCIETIKCNWATIDGVESKYGEORPSSSEFQDTOPSSDFEPP 1025

```
Qy      888 MADKLLLTTPSKFKTCQGPVDVNILAKNCPCLSNPCKNDGTCNSDPVDFYRCTCPYGFGK   947  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db      1 MKDKLILSTPSSFFVCGRVRNDILAKNCACFEQPCCNQQAVALPQREYQCQLCPQGYHG   60  
  
Qy     948 QDCDVPIHACISNPCKHGGTCHLKEGEEDGTWCICADGFEGENCEVNVDDC--EDNDCENN   1006  
: : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db     61 KHCEFMIDACYGNPCRNNAITCTVL--EEGRFSCQCAPGYTGARCETNIDDCGLGEIKQQNN   118  
  
Qy    1007 STCDVGINNYTCLCPPEYTGELCEEKLDFCAQDLNPCQHDSKCILLTPKGFCDCDCTPGYVG   1066  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    119 ATCIDGVESYKCECPQPGFSGEPCDTKIQFCSPFEFNPCANGAKCMDHFTHYSCDCQAGFHG   178  
  
Qy    1067 EHCIDIFFDDCNKCKNGAHCHTDAVNGYTCICPEGYSGLFCEFP--PMVLPRTPSPCDFN   1124  
: : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    179 TNCNTDNIDCCQNMCQGGTCDVGINDYQCRCPDDTYGKYCEGHNMISMYPQTSPCQNH   238  
  
Qy    1125 DCQNGA--QCIVRINEPICQLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTINITL   1182  
: : : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    239 ECKHGVCFPQNAQGS DYLCRHCGPYGTGWCEBYLTSISFVHNNSFVELEPLTRPEANVTI   298  
  
Qy    1183 QIATDEDSGILLYKGDKDHIAVELYRGVRASYDTGSHPASAIYSVETINDGNFHIVELL   1242  
: : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    299 VFSSGQ-NGIIIMYDGDQDAHLEVELFNGRIRVSYDVGNHPVSTMYSFEMVADGKYHAVELL   357  
  
Qy    1243 ALDQSLSLSDVGGNKPIITNLKSQSTLNFDSPLYVGGMPGKSNVASLRQAPQGNTSFGH   1302  
| : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    358 ATKKNFTLRVDRGLARSIINEGSNDYLKLTTPMFLGGLPVDAQAYKANWQIRNLTSFKG   417  
  
Qy    1303 CIRNLYINSELQDFQKVPMTGILGPCBCECHKVKVAHGTCPQSSQAGFTCECBQEWG--   1360  
: : : : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db    418 CMKEVVINHKLVDFGNAOROOKIPTGC-----ALLGEEOBBE-----DDEODFMDET   465
```

Qy 1503 TDGSSFFVDEVEKVVKCGC 1520
:|:: : | : :| |
Db 574 KNGTTKISTVHIROCOC 591

```

RESULT 9
A24420
notch protein - fruit fly (Drosophila melanogaster)
N;Alternate names: neurogenic repetitive locus protein
C;Species: Drosophila melanogaster
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A24420; A24768; S09358; A05267
R;Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6, 3094-3108, 1986
A;Reference number: A24420; MUID:87064624
A;Accession: A24420
A;Molecule type: DNA
A;Residues: 1-2703 <KID>
A;Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993
R;Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
Cell 43, 567-581, 1985
A;Reference number: A24768; MUID:86079539
A;Accession: A24768
A;Molecule type: mRNA
A;Residues: 1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',268-872,'R',874-9
A;Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2
R;Tautz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A;Title: Hypervariability of simple sequences as a general source for polymorphic DNA
A;Reference number: S09358; MUID:89385974
A;Accession: S09358
A;Molecule type: DNA
A;Residues: 2505-2551,'QQQQ',2552-2576,'E',2578-2604 <TAU>
R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985
A;Title: opa: a novel family of transcribed repeats shared by the Notch locus and
A;Reference number: A05267; MUID:85099329
A;Accession: A05267
A;Molecule type: DNA
A;Residues: 2504-2576,'E',2578-2611 <WHA2>
C;Genetics:
A;Gene: notch; opa
A;Cross-references: FlyBase:FBgn0004647
A;Map position: 8.96-9.36
A;Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: differentiation; tandem repeat; transmembrane protein
F;27-43/Domain: transmembrane #status predicted <TMM1>
F;297-328/Domain: EGF homology <EGX1>
F;530-561/Domain: EGF homology <EGF1>
F;568-599/Domain: EGF homology <EGF>
F;988-1019/Domain: EGF homology <EGX2>
F;1064-1095/Domain: EGF homology <EGF3>
F;1187-1218/Domain: EGF homology <EGX3>
F;1746-1762/Domain: transmembrane #status predicted <TMM2>
F;1950-1982/Domain: ankyrin repeat homology <AN1>
F;1983-2015/Domain: ankyrin repeat homology <AN2>
F;1988-2004/Domain: transmembrane #status predicted <TMM3>
F;2017-2049/Domain: ankyrin repeat homology <AN3>
F;2050-2082/Domain: ankyrin repeat homology <AN4>
F;2083-2115/Domain: ankyrin repeat homology <AN5>
F;2538-2568/Region: glutamine-rich
F;2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

```

Query Match : 9.8%; Score 818; DB 1; Length 2703;
Best Local Similarity 29.3%; Pred. No. 1.1e-39;
Matches 262; Conservative 93; Mismatches 336; Indels 202; Gaps 50;

Qy 707 TCDGNDNSCSPLSRCTECT---CLDTVVRCNSKGLKVLPGKIPRDVTLYLDGNQFT 763
||| | | : | | | | | : | | |
Db 307 TCIDGISDVTG---RCPPNETGRFCODDVDCARDHPVCONGATCTNTH----- 353

Db . 307 TCIDGISDYTC----RCPPNETGREGODVDDECAORDHPVCONGATCTNTH----- 353

DD: 307 TCIDGTSDFH---RCPFNFTGRFCQDDVDECAQRDAFVCQNGATCTNIA-----353

Qy 764 LVPKELSNYKHLTL-----IDLNNRISTLSNQSFNMTQLLLTILSYNRLRCLIPRPTFD 818
Db 354 -----GSYSICVNGWAGLDCSNNTDDCKQAACFYGAT-----CI-----D 389
Qy 819 GLKSLRLSLHGNDSVVEGAFNDLSALSHL--AIGANPLYCD--CNMQWLSL---WV 870
Db 390 GVGSFYCQCTKGK-----TGLLCHLDDACTSNPCHADAICTSPINGSYACSC 437
Qy 871 KSEYK---EPGIARC--AGPGEMADKLLLTTPSKKFTQ---GP-VDVNILAKCNPL 919
Db 438 ATGYKGVDCSEIDECQSGPCEHNGICVNTPGSYRNCSCQGTGPRCETNI---NECE 493
Qy 920 SNPCKNDGTGNSDPVDFYRCTCPYGFQDQCDVPIHACISNPCKHGGTCHLKEGEEDGF 979
Db 494 SHPCQNEGSLDDPGTF-RCVCMPTGTGTQCEIDECQSNPCLNDGTGCHDR---INGFK 549
Qy 980 CICADGFECEGCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGLCEELDFCAQD 1039
Db 550 CSCALGFTGARCQINIDDCQSQPCRNRGICHDSIAGYSCPEPPGTGTSCENINDC--D 607
Qy 1040 LNPQHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDQDNKCKNGAHTDAVNYTCICP 1099
608 SNPCHRGKCIDVNSFKCLCDPGYTGICQKQINECESNPQCFDGHQCDRVGSYCCQ 666
Qy 1100 EGYSGLCEFSPPMVLPRTPCDNFDQNGAQCIVRINEPICQCLPGYQGEKCEKLVSVN 1159
Db 667 AGTSGKNCEVN-----VNECHSNPCNMGATCIDGINSYKQCVPGTGTGCKE 714
Qy 1160 FINKESYLQIPSAKVRPQTNIQLIATDEDSGILLYKGDHIAVELYRGRVASYDTGS 1219
Db 715 --NVDECISSPCA-----NNGVCIDQVNG--YK-----CECPRG---FYD--A 748
Qy 1220 HPASAI---YSVETINDGNFHVALLDQSLSLSDVGGNPKII---TNLSKQSTLNF 1272
Db 749 HCLSDVDECAFNPCVNEGRCE-----DGNEFICHCPPTGKRCLELDID 793
Qy 1273 ---SPLYVGG-MPGKSNVASLRQAPQNG---TSFHGCI RNLYIN-----SELQDFQ- 1317
Db 794 ECSSNPCHGGTGYDKLNAFSCQCMPTGTGKCEINIDCVTNPCGNGGTGICDVNGYK 853
Qy 1318 --KVPMTGILPGCE---PCHKKVCAH-GTCQPSQ-AGFTCECQEGWMLCDQRTND 1369
Db 854 VCKVPTG--RDCEKMDPCASNCKNEAKTSSNFDLSCTCKLGYTGRYCDEDE 910
Qy 1370 PCLGNKCVHG-TCLPINAFSYCKCLEHGGVLCDEEDLFPQCAIKCKHKGCRSLG 1428
Db 911 CSLSSPCRNAGSCLNVPG-SYRCLCTKGYGRDCAINTD---DCASFPCCNGGTCLDG 966
Qy 1429 QPYCECSSGYTGDSCDREIS-----CR-GERIRDYQKQ---GYAA---CQTKKVS 1474
967 DYSCLVGVGDKKCECTDINECLSQPCNGATCSQYVNSYTCTPLGFSGINCQTND 1024
Qy 1475 RLEC-RGGCA-GG-CCGPLRSKRRKYSFECTDGSSFVDEVEKVVKCGCTRCVS 1525
Db 1025 --DCTESSCLNGGSCIDGING---YNCCLAGYSGANQYKLNKCDNSPCLN 1071

RESULT 10

A56136
jagged protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 11-Jan-2000
C:Accession: A56136
R:Lindsell, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.
Cell 80, 909-917, 1995
A:Title: Jagged: a mammalian ligand that activates Notch1.
A:Reference number: A56136; MUID:95211842
A:Accession: A56136
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1220 <LIN>
A:Cross-references: GB:L38483
C:Superfamily: unassigned EGF-related proteins; EGF homology

F:379-410/Domain: EGF homology <EGF1>
F:492-523/Domain: EGF homology <EGF>
F:634-665/Domain: EGF homology <EGF2>

Query Match 9.2%; Score 768; DB 2; Length 1220;
Best Local Similarity 25.4%; Pred. No. 3.4e-37;
Matches 197; Conservative 101; Mismatches 246; Indels 232; Gaps 32;

Qy 853 GANPLYCD-----CNMQWL---SDWKSEYKEPGIARCAGPGEMADKLLTT 896
Db 257 GWQGLYCKDCIPHPGCVHGTGCTNEPWCLETNW-----GGQLCKD----- 296
Qy 897 PSKFTQCGPVDVNILAKCNPLSNPCKNDGTGNSDPVDFYRCTCPYGFQDQCDVPIHA 956
Db 297 -----DLNYCGTHQCL-----NRGTCSTGPDYKQCSCPEGYSGPNCEIAEHA 340
Qy 957 CISNPCKHGGTCHLKEGEEDGFWICADGFECEGCEVNVDDCEDNDCENNSTCVDGINNY 1016
Db 341 CLSDPCHNRGSC--KE-TSSGFECECSPGWTGTCTSTNIDDCSPNNCSHGGTCDLVNGF 397
Qy 1017 TCLCPPEYTGLCEELDFCAQ-----DLN----- 1041
Db 398 KCVCPQWTGKTCQLDANECEAKPCVNARSKNLASYYCDCLPGWGMQCNIDINCLG 457
Qy 1042 PQCHDSKCILTPKGFKCDCTPGYVGEHCDIDFDDQDNKCKNGAHTDAVNYTCICPEG 1101
Db 458 QCNQDASCRDLVNGYRCICPPGYAGDHCEERDIDECASNPCNLNGHGCQNEINRFQCLCPTG 517
Qy 1102 YSGLCEFSPPMVLPRTPCDNFDQNGAQCIVRINEPICQCLPGYQGEKCEKLVSVNFI 1161
Db 518 FSGNLCQLD-----IDYCEPNPCQNGAQCYNRASDYFCCKPEDYEGKNCSHL----- 564
Qy 1162 NKESYLQIPSAKVRPQTNIQLIATDEDSGILLYKGDHIAVELYRGRVASYDTG--- 1218
Db 565 --KDHRTTPCEVIDSCTVAM-ASNDTPEGV-----RYISSNVCSPHCKCKSESQK 614
Qy 1219 ---SHPASAIYSVETINDGNFHVALLDQSLSLSDVGGNPN-----KIITNLSKQSTL 1269
Db 615 TCDCKNGFTGTGTCHENINDCE-----GNPCTNGTCTCIDGVNSYKCI 655
Qy 1270 NFDSPLYVGGMPG---KSNVASLRQAPQNGTSFHGCI RNLYIN----- 1310
Db 656 CSD-----GWEGAHCENNINDCSQNPCHYGGTCDRLVNDPYCDCKNGWKGTCHSRDSQ 709
Qy 1311 -----SELQDFQ-----KVPMTGILPGCEPCHKKVCAHGTQCP 1345
Db 710 CDEATCNNGGTCTYDEVDTFKCMCPGWEGTTCNIARNSSCLP--NPCHN---GGTCVVN 763
Qy 1346 SQAGFTCECQEGWMLCDQRTNDPCLGNKCVH-GTCLPINAFSYCKCLEHGGVLCDE 1404
Db 764 GDS-FTCVCKHGWEGPICTQNTND-CSPHPCYNSGTGVDGNW-YRCECAPGAGPDCRI 820
Qy 1405 EEDLFPNCAIKCKHKGKRLSGLQPYCECSSGYTGDSCDREISCR-----GERIRDYQ 1459
Db 821 N---INEQSSPCAFGATCVDEINGYQCICPPGHSGAKC-HEVSGRSCITMGRVILDGAK 876
Qy 1460 KQGYAACQTKKVSRLCEKRGCCAGGQ--C---CGPLRSKRRKYSFECTDGSSFV 1509
Db 877 WDDDCNTCQ-----CLNGRVACSKVWCGPRPCRLKHGHECPNGQSCI 919

RESULT 11

A35844
Xotch protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35844
R:Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A:Title: Xotch, the Xenopus homolog of Drosophila notch.
A:Reference number: A35844; MUID:90385285
A:Accession: A35844

Db 633 TTGVNCE TKID---DCASNLCDNGKCIDKIDGY---ECTCEPGYTGKLCNININECD SNP 686

Dh 543 GTCTDEI-----GYYTCTCPTGTGSSCEINPDDCVGNPCOY-GTC-VDGVDDYSCSCT 594

```
Qy 861 RCAGPGEMADKLLLTTPSKK-----FTC-----QGPDVNILAKNCNPLS 920
   || | :||| : | ||: : || ||:
Db 56 RQCDPSP-----CLSTPCKNAGTCYVVDHGGIVDASCPLGFGPLCLTALA--NACLA 108

Qy 921 NPCKNDGTGN-----SDPV-----DFYRCTCPY 943
   |||:| |||: :|| : |||
Db 109 NPCNRGGTCDLTLLEYKCRCPPWGSKGSCQQADPCASNPCANGGQLPFESSYICGCPP 168

Qy 944 GFKGQDCDVPYHACISNP--CKHGCTHLKEGE----- 974
   || | | : : | | |:|||| : |
Db 169 GFHGPTCRQDVNECSQNPLGRHGGTCHNEIGSYRCACRATHTGPHCELPHYVPCSPSPCQ 228

Qy 975 -----EDGFWCICADGFEGENECVNVDDCEDNCCENNSTCVDGINNYTCLCPPEYT 1025
   | | | | |:||| |||| |:|:| |||:| | | |||:|
Db 229 NGGTCRPTGDTHETACALPGFAGQCNCEENVDDCPGNCKNGGACVDGVNTINCRCPPWT 288

Qy 1026 GELCEEKLD----- 1034
   |:| |:|
Db 289 GQYCTEDVDEQCMLPNACQNAGTCHNSHGGYNVCVNGWTGEDCSDNIDDCASAAFCQA 348

Qy 1035 -----F 1035

Db 349 TCHDRVASFYCECPHGRITGLLCHLNDAICSNPCNEGSNCDTPNVNGKAICTCPRGYTGA 408

Qy 1036 CAQDL-----NPCQHDSKCILTPKGFKCDCTPGYVGHECDIFDDCQDNKCKNGAHT 1088
   |:|:| |||:| ||:| |:| |:| |:| |:| |:| |:| |:|
Db 409 CSQDVDECALGANPECAGKHCNLWLTSFPECQCLQGYTGPRCEIDVNECISNQCNDATCL 468

Qy 1089 DAVNGYTCICPEGYSGLFCEFSPPMVLPRTPSPDNFDQNAQOCIVRINEPICQLPGYQ 1148
   | : : |||:| ||::|| : | : : : : : : : : : : : : : :
Db 469 DQIEGFQCI CMPEGYEGVICEIN-----TDECASSPLHNRCRVKDINEFLQCCKGFS 521

Qy 1149 GEKC---EKLVSVNFINKESYLQIPSAKVRPQTNITLQIATEDSGILLYKGDKDHIAY 1204
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 522 GHLCQYDVDECASTPCKNAGKCLDGPNT----YTCVCTEGYTGTHCEVIDECDOPDPCHI 577

Qy 1205 ELYGRVRASTDYGSHPASAIYSVET-IND-----GNFHIVELLA----- 1243
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 578 GLCKDGV-ATFTCLCQPGYTGHHCETNINECHSQPCRHGTCQDRNNYLLCLCLKGTGP 636

Qy 1244 ----LDQSLSLVDDGNPRIITNLKSQSTLNFD-----SPLYVGMPGKSNVASLRQA 1292
   | | | | | | | | | | | | | | | | | | | | | |
Db 637 NCEINLDDCASNPDCSG----TCLDK----IDGYECACEPGYTGSM-CMVNIDECAGS 685

Qy 1293 PGQNGTSFHGCIRNLYINSELQDFQVPMQTGIL-----PGC----EPCHKKV 1336
   | | | : | : | : | : | : | : | : | : | : | : |
Db 686 PCHNGGT-----CEDGIAGTFCRCPEGYHDPTCLSEVNECSNP 724

Qy 1337 CAHGTCPQSSAQGFTCECBQBGWMGPLCDQRNTDPLGNKCVH-GTCLPINAFSYSCKCLE 1395
   | | | : : : : | : | : | : | : | : | : | : | : | : |
Db 725 CIHGACRDGLN-GYKCDAPGWSGTNCDINNNE-CESNPCVNGGTCKDMTS-GYVCTCRE 781

Qy 1396 HGHHVLCDEE-----EDLFNCPQAIKKKH- 1419
   | | | | | | | | | | | | | | | | | |
Db 782 FGSGPNCQTNINECASNPCLNQGTICDDVAGYKCNCLPYTGATCEVVLAPCATSPCKNS 841

Qy 1420 GKRLSGSLQOPY-CECSSGYTTGSDCREIS-----CRGERIRDYYQKQGYAACQTKK 1472
   | | | : : : : | : | : | : | : | : | : | : | : | : |
Db 842 GVCKESEDYESFSCVPTGWQQQTCEIDINECVKSPCR-----HG-ASCQNTNG 889

Qy 1473 VSRLECRGGCGAGQC-----CGPLRSKRKRKYSFECTDG--SSFVDEV-----EKV 1515
   | | : | | | | | | | | | | | | | | | | |
Db 890 SYRCLQAGYTGRCNESDIDCRPNPCHN---GGSCTDGVNAAFCDLPGFQGAFCEDI 946

Qy 1516 VKC-----GCTRCV 1524
   : | : | | |
Db 947 NECATNPCONGANCTDCV 964
```

F;2050-2082/Domain: ankyrin repeat homology <AN5>

```

Qy      662 FNCNCYLAWLGEWLKKRIVTGNPRCKPYFLKEIQDVAIQDFTCDDGNDNDSNCSPLS 721
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      202 YRCACCATHTG-----PHCELPY-----VPCSPSPCQNG--ATCRPTG 237

Qy      722 RCPTECTCLDVTV-----RCSNGLKVLPLKGI-----PRDVTLEYL-D 758
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      238 DTHTEACLPGFAGONCEENVDDCPGNMCKNGGACV--DGVNTYRCRCPPTVGOYCTED 295

```

[illegible]

```

RESULT 15
A49175
Motch B protein - mouse (fragment)
N;Alternate names: Notch homolog
C;Species: Mus musculus (house mouse)
C;Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
C;Accession: A49175; PH1570; S32113
R;Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A;Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues
A;Reference number: A49175; MUID:93178563

```

Search completed: January 22, 2001, 12:21:29
Job time: 1766 sec

Search completed: January 22, 2001, 12:21:29
Job time: 1766 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:08:19 ; Search time 162.41 Seconds
(without alignments)
303.236 Million cell updates/sec

Title: US-09-540-245A-2
Perfect score: 8316
Sequence: 1 MRGVGVNQMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3486	41.9	1480	1	SLIT_DROME	P24014 drosophila
2	818	9.8	2703	1	NOTC_DROME	P07207 drosophila
3	759	9.1	2524	1	NOTC_XENLA	P21783 xenopus lae
4	747.5	9.0	2531	1	NTC1_RAT	Q07008 rattus norv
5	745	9.0	2531	1	NTC1_MOUSE	Q01705 mus musculu
6	735.5	8.8	2444	1	NTC1_HUMAN	P46531 homo sapien
7	732.5	8.8	2437	1	NOTC_BRARE	P46530 brachydanio
8	731	8.8	1064	1	FBP1_STRPU	P10079 strongyloce
9	722	8.7	2318	1	NTC3_MOUSE	Q61982 mus musculu
10	691.5	8.3	1964	1	NTC4_MOUSE	P31695 mus musculu
11	634.5	7.6	2139	1	CRB_DROME	P10040 drosophila
12	625.5	7.5	1408	1	SERR_DROME	P18168 drosophila
13	606	7.3	1429	1	LI12_CAEEL	P14585 caenorhabdi
14	549.5	6.6	570	1	FBP3_STRPU	P49013 strongyloce
15	543	6.5	1295	1	GLP1_CAEEL	P13508 caenorhabdi
16	535.5	6.4	473	1	FP2_MYTGA	Q25464 mytilus gal
17	533.5	6.4	723	1	DLL1_HUMAN	O00548 homo sapien
18	521.5	6.3	714	1	DLL1_RAT	P97677 rattus norv
19	518.5	6.2	722	1	DLL1_MOUSE	Q61483 mus musculu
20	493.5	5.9	603	1	ALS_MOUSE	P70389 mus musculu
21	491	5.9	833	1	DL_DROME	P10041 drosophila
22	489	5.9	603	1	ALS_RAT	P35859 rattus norv
23	484.5	5.8	605	1	ALS_HUMAN	P35858 homo sapien
24	479.5	5.8	605	1	ALS_PAPPA	O02833 papio papio
25	455.5	5.5	1134	1	CHAO_DROME	P12024 drosophila
26	442.5	5.3	3051	1	YXK3_CAEEL	P34576 caenorhabdi
27	410	4.9	567	1	GPV_MOUSE	O08742 mus musculu
28	403.5	4.9	5147	1	FAT_DROME	P33450 drosophila
29	400	4.8	567	1	GPV_RAT	O08770 rattus norv
30	400	4.8	2871	1	FBN1_BOVIN	P98133 bos taurus
31	397	4.8	2871	1	FBN1_HUMAN	P35555 homo sapien
32	393.5	4.7	2871	1	FBN1_MOUSE	Q61554 mus musculu
33	393	4.7	3707	1	PGBM_MOUSE	Q05793 mus musculu

34	390.5	4.7	4393	1	PGBM_HUMAN	P98160 homo sapien
35	390	4.7	662	1	GARP_HUMAN	Q14392 homo sapien
36	389.5	4.7	2911	1	FBN2_HUMAN	P35556 homo sapien
37	387.5	4.7	1959	1	AGRI_RAT	P25304 rattus norv
38	386.5	4.6	2907	1	FBN2_MOUSE	Q61555 mus musculu
39	382.5	4.6	560	1	GPV_HUMAN	P40197 homo sapien
40	382	4.6	4590	1	FATH_HUMAN	Q14517 homo sapien
41	376.5	4.5	1097	1	TOLL_DROME	P08953 drosophila
42	375.5	4.5	385	1	DLK_MOUSE	Q09163 mus musculu
43	372	4.5	383	1	DLK_HUMAN	P80370 homo sapien
44	372	4.5	536	1	CBP8_HUMAN	P22792 homo sapien
45	371	4.5	4289	1	TENX_HUMAN	P22105 homo sapien

ALIGNMENTS

RESULT 1
SLIT_DROME
ID SLIT_DROME STANDARD; PRT; 1480 AA.
AC P24014;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SLIT PROTEIN PRECURSOR.
GN SLI.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91099665; PubMed-2176636;
RA Rothberg J.M., Jacobs J.R., Goodman C.S., Artavanis-Tsakonas S.;
RT "Slit: an extracellular protein necessary for development of midline
RT glia and commissural axon pathways contains both EGF and LRR
RT domains.";
RL Genes Dev. 4:2169-2187(1990).
CC -1- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC MATRIX MOLECULES.
CC -1- ALTERNATIVE PRODUCTS: GIVES RISE TO 2 DISTINCT PROTEINS DIFFERING
CC BY 11 AA AT THE C-TERMINUS OF THE LAST EGF REPEAT.
CC -1- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC -1- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 22. TWO BLOCK OF 6 LRR'S
CC AND TWO BLOCKS OF 5 LRR'S.
CC -1- SIMILARITY: CONTAINS A C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X53959; CAA37910.1; -.
DR PIR; A36665; A36665.
DR HSP; P00740; 1XA.
DR FLYBASE; FBgn0003425; sli.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00007; Cys_knot; 1.
DR PFAM; PF00008; EGF; 7.

DR	PFAM; PF00560; LRR; 16.	
DR	PFAM; PF01463; LRRCT; 4.	
DR	PFAM; PF01462; LRRNT; 4.	
DR	PFAM; PF00054; laminin_G; 1.	
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.	
DR	PROSITE; PS00022; EGF_1; 7.	
DR	PROSITE; PS01185; CTCK_1; 1.	
DR	PROSITE; PS01186; EGF_2; 5.	
DR	PROSITE; PS01187; EGF_CA; 2.	
DR	PROSITE; PS01225; CTCK_2; 1.	
KW	Neurogenesis; Glycoprotein; Signal; Alternative splicing;	
KW	EGF-like domain; Repeat; Leucine-repeat; Duplication.	
FT	SIGNAL	1 36
FT	QHAIN	37 1480
FT	DOMAIN	70 104
FT	DOMAIN	105 230
FT	DOMAIN	231 294
FT	DOMAIN	295 326
FT	DOMAIN	327 452
FT	DOMAIN	453 518
FT	DOMAIN	519 550
FT	DOMAIN	551 653
FT	DOMAIN	654 714
FT	DOMAIN	715 746
FT	DOMAIN	747 848
FT	DOMAIN	849 910
FT	REPEAT	105 115
FT	REPEAT	116 139
FT	REPEAT	140 163
FT	REPEAT	164 187
FT	REPEAT	188 211
FT	REPEAT	212 230
FT	REPEAT	327 337
FT	REPEAT	338 361
FT	REPEAT	362 385
FT	REPEAT	386 409
FT	REPEAT	410 433
FT	REPEAT	434 452
FT	REPEAT	551 562
FT	REPEAT	563 586
FT	REPEAT	587 610
FT	REPEAT	611 634
FT	REPEAT	635 653
FT	REPEAT	747 757
FT	REPEAT	758 781
FT	REPEAT	782 805
FT	REPEAT	806 829
FT	REPEAT	830 848
FT	DOMAIN	907 944
FT	DOMAIN	946 983
FT	DOMAIN	985 1022
FT	DOMAIN	1024 1062
FT	DOMAIN	1064 1100
FT	DOMAIN	1111 1149
FT	DOMAIN	1353 1392
FT	DOMAIN	1409 1480
FT	CARBOHYD	111 111
FT	CARBOHYD	207 207
FT	CARBOHYD	357 357
FT	CARBOHYD	435 435
FT	CARBOHYD	783 783
FT	CARBOHYD	788 788
FT	CARBOHYD	958 958
FT	CARBOHYD	998 998
FT	CARBOHYD	1060 1060
FT	CARBOHYD	1159 1159
FT	CARBOHYD	1175 1175
FT	CARBOHYD	1243 1243
FT	CARBOHYD	1292 1292
FT	DISULFID	911 922
FT	DISULFID	916 932
FT	DISULFID	934 943
FT	DISULFID	950 961
		SLIT PROTEIN.
		CONSERVED N-FLANKING REGION OF THE LRR.
		LEUCINE-RICH REPEATS (1ST REGION).
		CONSERVED C-FLANKING REGION OF THE LRR.
		CONSERVED N-FLANKING REGION OF THE LRR.
		LEUCINE-RICH REPEATS (2ND REGION).
		CONSERVED C-FLANKING REGION OF THE LRR.
		CONSERVED N-FLANKING REGION OF THE LRR.
		LEUCINE-RICH REPEATS (3RD REGION).
		CONSERVED C-FLANKING REGION OF THE LRR.
		CONSERVED N-FLANKING REGION OF THE LRR.
		LEUCINE-RICH REPEATS (4TH REGION).
		CONSERVED C-FLANKING REGION OF THE LRR.
		LRR 1-1.
		LRR 1-2.
		LRR 1-3.
		LRR 1-4.
		LRR 1-5.
		LRR 1-6.
		LRR 2-1.
		LRR 2-2.
		LRR 2-3.
		LRR 2-4.
		LRR 2-5.
		LRR 2-6.
		LRR 3-1.
		LRR 3-2.
		LRR 3-3.
		LRR 3-4.
		LRR 3-5.
		LRR 4-1.
		LRR 4-2.
		LRR 4-3.
		LRR 4-4.
		LRR 4-5.
		EGF-LIKE 1.
		EGF-LIKE 2.
		EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
		EGF-LIKE 4.
		EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
		EGF-LIKE 6.
		EGF-LIKE 7.
		CTCK.
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		N-LINKED (GLCNAC. . .) (POTENTIAL).
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.
		BY SIMILARITY.

FT	DISULFID	955	971	BY SIMILARITY.
FT	DISULFID	973	982	BY SIMILARITY.
FT	DISULFID	989	1001	BY SIMILARITY.
FT	DISULFID	995	1010	BY SIMILARITY.
FT	DISULFID	1012	1021	BY SIMILARITY.
FT	DISULFID	1028	1041	BY SIMILARITY.
FT	DISULFID	1035	1050	BY SIMILARITY.
FT	DISULFID	1052	1061	BY SIMILARITY.
FT	DISULFID	1068	1079	BY SIMILARITY.
FT	DISULFID	1073	1088	BY SIMILARITY.
FT	DISULFID	1090	1099	BY SIMILARITY.
FT	DISULFID	1115	1125	BY SIMILARITY.
FT	DISULFID	1120	1137	BY SIMILARITY.
FT	DISULFID	1139	1148	BY SIMILARITY.
FT	DISULFID	1357	1368	BY SIMILARITY.
FT	DISULFID	1362	1380	BY SIMILARITY.
FT	DISULFID	1382	1391	BY SIMILARITY.
FT	DISULFID	1409	1443	BY SIMILARITY.
FT	DISULFID	1423	1457	BY SIMILARITY.
FT	DISULFID	1434	1473	BY SIMILARITY.
FT	DISULFID	1438	1475	BY SIMILARITY.
FT	DISULFID	1442	1479	BY SIMILARITY.
FT	VARSPLIC	1394	1404	MISSING (IN SHORT ISOFORM).
SO	SEQUENCE	1480 AA:	165752 MW:	F9D5925FC170B1C3 CRC64:

Query Match 41.9%; Score 3486; DB 1; Length 1480;
Best Local Similarity 43.8%; Pred. No. 1.5e-210;
Matches 660; Conservative 275; Mismatches 457; Indels 116; Gaps 22;

[illegible]

Qy 686 RCQKPYFLKEIPQDVAIQDFTCDGNDNSCSPLSRCTECTCLDTVVRCNSKGLKVP 745
| | : : : : | : | : | : | | | | | | : :
Db 679 RCGAPSKVRDQIKDLPHSEFKCSSENSE-GCLGDGVCPPSCTCTGTVACSRNQLKEIP 737

Qy 746 KGIPRDVTELYLDGNQFTLVPE-LSNYKHLTLIDLSNNRISTLSNQSFNNMTQLTLIL 804
: | | : : | | : : | : : : | : | | : : | : | : | : | :
Db 738 RGIPAEITSELYLESNEIEQIHYERIRHLRSLTRLDLSNNQITILSNYTFANLTKLSTLII 797

Qy 805 SYNLRCPPTFDGLKSLRLSLHGNDSIVVPEGAFNDLSALSHLAGANPLYCDNCMQ 864
| | : : : : | : | : | | | : | : | : | : | : | : | :
Db 798 SYNLQCLQRHALSGLNNLRVVSILHGNRISMLPEGSFEDLSLTHIALGSNPLYCDGCLK 857

Qy 865 WLSDWVSEYKEPGIARCAGPEMADKLLLTTPSKKFTCGPVDVNILAKNCPCLSNPK 924
| | | : | : | | | | | : | : | : | : | : | : | : | : | :
Db 858 WFSWIKLDLYVEPGIARCAEPEQMKKLLISTPSSSVFCGRVRNDILAKNACFEQPCQ 917

925 NDGTCNSDPVDFRCTCPYGFQGDQDVIHACISNPKHGCTCHLKEGEEDGFWICAD 984
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 918 NQAQCVLPQREYQCLCPGYHGRKHECFMIDACYNPCNNATCTVL--EGRFSCQCAP 975

Qy 985 GFEGENCEVNVDDC-EDNDCENNSTCVDGINNYTCLCPPEYTGECEKLOFCAQDLNCP 1043
| : | | | : | : | : | : | : | : | : | : | : | : | : | :
Db 976 GTTGARCTNIDDCLEIKCQNNATCIDGVSEKCECPGSGFCDTKIQCFSPFENPC 1035

Qy 1044 QHDSKCLTPKGFKCDCTPGYVGEHCDIDFDDQDNKCKNAHCTDAVNGYTCICPEGYS 1103
: | | : : | : | : | : | : | : | : | : | : | : | : | :
Db 1036 ANGAKMDHFTYSCDCQAGFHGTNCTNIDDCQNHMCNGGTCVDGINDYQCRCPDDYT 1095

Qy 1104 GLFCEPSP--PMVLPRTSPCDNFCQNGA--QCIVRINEPICQLPGYQGECEKLVSN 1159
| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1096 GKYCEGHNMISMYPQSPCNHCEKHHVCFQPNAGSDYLCRCHPGYTGKVEYLSIS 1155

Qy 1160 FINKESYLQIPSAKVRPQTNIQLIATDEDSGILLYKGDHIAVELYRGRVASYDTGS 1219
| : | : : : | : | : | : | : | : | : | : | : | : | : | :
Db 1156 FVHNNSFVELEPLRTRPEANVTIWFSSAQNGILMYDQDAHLAVELNGRIVSYDVGN 1215

Qy 1220 HPASAIYSVETINDGNFHVILLALDQSLSLSDVGGNPKIITNLKQSTLNFDSPLYVG 1279
| | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1216 HPVSTMYSEFEMVADGKYHAVELLAIAKNFTLRVDRGLARSIIINEGSNDYIKLTPMPLGG 1275

Qy 1280 MPKSNVASLRQAPGQNGTSFHGCIIRNLYINSELQDFQKVPMTGLPGCEPCHKVCAH 1339
: | : : : : | : | : | : | : | : | : | : | : | : | :
Db 1276 LPVDPAQAYKNWQIRNLTSFGKCKMEVWNNHKLVDPGNAQRQKITPGC-----ALLE 1329

Qy 1340 GTCQPSQAGFTCEQEGWGWG--PLCDQRTNDPCLGNKCVHGT-CLP-INA-FSYCKCL 1394
| | : : : : | : | : | : | : | : | : | : | : | : | :
Db 1330 GEQEQEE-----DDEQDFMDETPHIKEEPVDPCLENKCRGSRCPNSNARDGYQCKCK 1383

Qy 1395 EGHGGVLCDEEEDLNPCQAIKCKHKGKRLSLGSLGQYPCBSSGYTGDSCDREISCRGERI 1454
| | | : | : | : | : | : | : | : | : | : | : | : | :
Db 1384 HGQGRGYCDQGEGETEP-----PVTAA-----TCRKEQV 1414

Qy 1455 RDYQYQKQGYAACQTTKKVSRLECRGGCAGGCGGCLPLRSKRRKYSPECTDGSSEFVDEVEK 1514
| : | : | : : : | : | | | : | : | : | : | : | : : :
Db 1415 REYTTEND----CRSROPKYAKCVGGC-GNCCAARIVRRKVRMVCNNRKYIKNLDI 1469

Qy 1515 VVKCGCTR 1522
| | | | : :
Db 1470 VRKCGCTK 1477

RESULT 2

NOTC_DROME

ID NOTC_DROME STANDARD; PRT: 2703 AA.
AC P07207; P04154;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN N.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats.";
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors.";
RL Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE OF 2505-2611 FROM N.A.
RX MEDLINE=85099329; PubMed=2981631;
RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
RT "opa: a novel family of transcribed repeats shared by the Notch locus
RT and other developmentally regulated loci in D. melanogaster.";
RL Cell 40:55-62(1985).
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE=87257846; PubMed=3037327;
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
RT "Restriction of P-element insertions at the Notch locus of Drosophila
RT melanogaster.";
RL Mol. Cell. Biol. 7:1545-1548(1987).
RN [5]
RP REVIEW.
RA Harris W.A.;
RT "Many cell types specified by Notch function.";
RL Curr. Biol. 1:120-122(1991).
CC -1- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC ECTODERM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
CC NEUROGENIC GENES.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16152; AAB59220.1; -.
DR EMBL; M16153; AAB59220.1; JOINED.
DR EMBL; M16149; AAB59220.1; JOINED.
DR EMBL; M16150; AAB59220.1; JOINED.
DR EMBL; M16151; AAB59220.1; JOINED.
DR EMBL; K03508; AAA28725.1; -.
DR EMBL; M13689; AAA28725.1; JOINED.
DR EMBL; K03507; AAA28725.1; JOINED.
DR EMBL; M12175; AAA74496.1; -.
DR EMBL; M16025; AAA28726.1; -.
DR PIR; A24420; A24420.
DR PIR; A24768; A24768.
DR PIR; A05267; A05267.
DR HSSP; P00740; 1IXA.
DR FLYBASE; FBgn0004647; N.
DR INTERPRO; IPR000152; -.

DR	INTERPRO; IPR000561; -.		
DR	INTERPRO; IPR000800; -.		
DR	INTERPRO; IPR001438; -.		
DR	INTERPRO; IPR001881; -.		
DR	INTERPRO; IPR002110; -.		
DR	PFAM; PF00008; EGF; 36.		
DR	PFAM; PF00023; ank; 6.		
DR	PFAM; PF00066; notch; 3.		
DR	PRINTS; PR00010; EGFBL00D.		
DR	PROSITE; PS50088; ANK_REPEAT; 5.		
DR	PROSITE; PS50297; ANK_REP_REGION; 1.		
DR	PROSITE; PS00010; ASX_HYDROXYL; 22.		
DR	PROSITE; PS00022; EGF_1; 34.		
DR	PROSITE; PS01186; EGF_2; 28.		
DR	PROSITE; PS01187; EGF_CA; 22.		
KW	Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;		
KW	Transmembrane; Signal; Glycoprotein.		
FT	SIGNAL	1 44	POTENTIAL.
FT	CHAIN	45 2703	NEUROGENIC LOCUS NOTCH PROTEIN.
FT	DOMAIN	45 1745	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1746 1766	POTENTIAL.
FT	DOMAIN	1767 2703	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	58 1451	36 X EGF-TYPE REPEATS.
FT	DOMAIN	58 95	EGF-LIKE 1.
FT	DOMAIN	96 136	EGF-LIKE 2.
FT	DOMAIN	139 176	EGF-LIKE 3.
FT	DOMAIN	177 215	EGF-LIKE 4.
FT	DOMAIN	217 253	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	255 291	EGF-LIKE 6.
FT	DOMAIN	293 329	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	331 370	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372 408	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	409 447	EGF-LIKE 10.
FT	DOMAIN	449 486	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	488 524	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	526 562	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	564 600	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	602 637	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	639 675	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	677 713	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	715 751	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	753 789	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	791 827	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	829 865	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	867 905	EGF-LIKE 22.
FT	DOMAIN	907 944	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	946 982	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	984 1020	EGF-LIKE 25.
FT	DOMAIN	1022 1058	EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1060 1096	EGF-LIKE 27.
FT	DOMAIN	1098 1134	EGF-LIKE 28.
FT	DOMAIN	1136 1181	EGF-LIKE 29.
FT	DOMAIN	1183 1219	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1221 1257	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1259 1295	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1297 1335	EGF-LIKE 33.
FT	DOMAIN	1337 1373	EGF-LIKE 34.
FT	DOMAIN	1375 1412	EGF-LIKE 35.
FT	DOMAIN	1415 1451	EGF-LIKE 36.
FT	DOMAIN	1475 1593	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1475 1513	LIN/NOTCH 1.
FT	REPEAT	1514 1553	LIN/NOTCH 2.
FT	REPEAT	1554 1593	LIN/NOTCH 3.
FT	DOMAIN	1896 2109	6 X ANK MOTIF REPEATS.
FT	DOMAIN	2538 2568	POLY-GLN (OPA-REPEAT).
FT	DISULFID	62 73	BY SIMILARITY.
FT	DISULFID	67 83	BY SIMILARITY.
FT	DISULFID	85 94	BY SIMILARITY.
FT	DISULFID	100 111	BY SIMILARITY.
FT	DISULFID	105 124	BY SIMILARITY.
FT	DISULFID	126 135	BY SIMILARITY.
FT	DISULFID	143 154	BY SIMILARITY.
FT	DISULFID	148 164	BY SIMILARITY.

FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	181	192	BY SIMILARITY.
FT	DISULFID	186	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	241	BY SIMILARITY.
FT	DISULFID	243	252	BY SIMILARITY.
FT	DISULFID	259	270	BY SIMILARITY.
FT	DISULFID	264	279	BY SIMILARITY.
FT	DISULFID	281	290	BY SIMILARITY.
FT	DISULFID	297	308	BY SIMILARITY.
FT	DISULFID	302	317	BY SIMILARITY.
FT	DISULFID	319	328	BY SIMILARITY.
FT	DISULFID	335	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	376	387	BY SIMILARITY.
FT	DISULFID	381	396	BY SIMILARITY.
FT	DISULFID	398	407	BY SIMILARITY.
FT	DISULFID	413	424	BY SIMILARITY.
FT	DISULFID	418	435	BY SIMILARITY.
FT	DISULFID	437	446	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	459	474	BY SIMILARITY.
FT	DISULFID	476	485	BY SIMILARITY.
FT	DISULFID	492	503	BY SIMILARITY.
FT	DISULFID	497	512	BY SIMILARITY.
FT	DISULFID	514	523	BY SIMILARITY.
FT	DISULFID	530	541	BY SIMILARITY.
FT	DISULFID	535	550	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	568	579	BY SIMILARITY.
FT	DISULFID	573	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	616	BY SIMILARITY.
FT	DISULFID	611	625	BY SIMILARITY.
FT	DISULFID	627	636	BY SIMILARITY.
FT	DISULFID	643	654	BY SIMILARITY.
FT	DISULFID	648	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	730	BY SIMILARITY.
FT	DISULFID	724	739	BY SIMILARITY.

Query Match 9.8%; Score 818; DB 1; Length 2703;
Best Local Similarity 29.3%; Pred. NO. 3.7e-43;
Matches 262; Conservative 93; Mismatches 336; Indels 202; Gaps 50;

```

Qy      707 TCDDGNDNNSCSPLSRCPTECT---CLDTVVRCSNKGKLVKPGKIPRVDTELYLDGNQFT 763
      || || | : | || | | | : : | | |
Db      307 TCIDIGISDYTC---RCPPNFTGRFCQDDVDECAQRDHPVCQNGATCTNTH----- 353

Qy      764 LVPKELSNYKHLTL-----IDLNNRISTLSNQSFNMTQLTLILSYNRLRCIPPTFD 818
      : | : : : | || | | | | | | | |
Db      354 -----GSYSICVNGWAGLDCSNNTDCKQAACFYGAT-----CI-----D 389

Qy      819 GLKSLRLLSLHGNDISVVEGAFNDLSALSHL--AIGANPLYCD--CNMQWLSD----WV 870
      | : | | | | | | | | | | | | | | : :
Db      390 GVGSFYQCCTKGK-----TGLLCHLDDACTSNPCHADAICDTSPIINGSYACSC 437

Qy      871 KSEYK---EPGIARC--AGPGEMADKLLTTPSKKFTCQ---GP-VDVNILAKCNPL 919
      : | | | | | | | | | | | | | | |
Db      438 ATGYKGVDCESEDIDECQSPCEHNGICVNTPGSYRCNCSQGGTGPRCETNI---NECE 493

Qy      920 SNPCKNDGTCNSDPVDFYRCTCPYFGKQDCQDVIHACISNPCKHGGTCHLKEGEEDGF 979
      | : | : | : | | | | | | | | : | : | | | : |
Db      494 SHPCQNEGSCILDGPGTF-RCVCMPGFTGTQCEIDIDEQCSNPCLNDGTCHDK---INGFK 549

Qy      980 CICADGFEGENCEVNVDDCEDNDCCNNSTCVDGINNYTCLCPPEYTGELCEEKLDFCQA 1039
      : | : | : | : | : | : | : | : | : | : | : |
Db      550 CSCALGETGARCOINIDDCOSOPCRNRGICHDSIAGYSCRCPPGYTGTSCSEININD--D 607

```

```

RESULT 3
NOTC_XENLA
ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
XOTCH.
Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M33874; AAB02039.1; -.
DR PIR; A35844; A35844.
DR HSSP; P00740; 1IXA.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1729 1750 POTENTIAL.
FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 57 EGF-LIKE 1.
FT DOMAIN 58 99 EGF-LIKE 2.
FT DOMAIN 102 140 EGF-LIKE 3.
FT DOMAIN 141 177 EGF-LIKE 4.
FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 489 525 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 527 563 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 565 600 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 602 638 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 640 675 EGF-LIKE 17.
FT DOMAIN 677 713 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 715 750 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 752 788 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 790 826 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 828 866 EGF-LIKE 22.
FT DOMAIN 868 904 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 906 942 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 944 980 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 982 1018 EGF-LIKE 26.
FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1058 1094 EGF-LIKE 28.
FT DOMAIN 1096 1142 EGF-LIKE 29.
FT DOMAIN 1144 1180 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1266 1304 EGF-LIKE 33.
FT DOMAIN 1306 1346 EGF-LIKE 34.
FT DOMAIN 1347 1383 EGF-LIKE 35.
FT DOMAIN 1386 1424 EGF-LIKE 36.
FT DOMAIN 1441 1560 3 X LIN/NOTCH REPEATS.
FT REPEAT 1441 1478 LIN/NOTCH 1.
FT REPEAT 1479 1520 LIN/NOTCH 2.
FT REPEAT 1521 1560 LIN/NOTCH 3.
FT DOMAIN 1871 2083 6 X ANK MOTIF REPEATS.
FT DISULFID 22 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.

```


QY 1315 DFQKVPMTGILPGCEPCHKKVCAHGTCQPSQAGFTCECQEGWMPGLCDQRTND----P 1370
Db 718 E-----CNSNPCIHGACHDGVN-GYKCDCEAGWSGNCNDINNECESNP 760

QY 1371 CL-----GNKCV-HGTCLPINFASYSCKLEG 1396
Db 761 CMNGGTCKDMTGAYICTCKAGFSGPNCQTINECSSNCLNHTGTCID-DVAGYKCNCLMP 819

QY 1397 HGGVLCDEEDLFNCPQAIKCKH-GKCRSLGGLQPY-CECSSGYTGDSCDREIS----- 1448
Db 820 YTGAIK---EAVLAPCAGSPCKNGGRCKESEDFTFSCBPPGQGTCEIDMNECVNRP 876

QY 1449 CRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGQC-----CGPLSRKRRKYSPEC 1502
Db 877 CRNG-----ATCQNTNGSYKCNCKPGYTGRCNEMDIDDCQP-----NPC 915

1503 TDGSSVFDEV-----EKVVK-----GCTRCVS 1525
916 HNGGSCSDGINMFFCNCAGFRGPKCEEDINECASNPCKNGANCTDCVN 964

RESULT 4

NTC1_RAT

ID NTC1_RAT STANDARD; PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SCHWANN CELL;
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RT development.";
RL Development 113:199-205(1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
OF TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC ADULT.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57405; CAA40667.1; -.
DR HSPS; P00740; 1122.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002049; -.

DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGFAMININ.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1724 1746 POTENTIAL.
FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 EGF-LIKE 1.
FT DOMAIN 59 99 EGF-LIKE 2.
FT DOMAIN 102 139 EGF-LIKE 3.
FT DOMAIN 140 176 EGF-LIKE 4.
FT DOMAIN 178 216 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 218 255 EGF-LIKE 6.
FT DOMAIN 257 293 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 295 333 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 335 371 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 410 EGF-LIKE 10.
FT DOMAIN 412 450 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 716 751 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22.
FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24.
FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26.
FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28.
FT DOMAIN 1097 1143 EGF-LIKE 29.
FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33.
FT DOMAIN 1307 1346 EGF-LIKE 34.
FT DOMAIN 1348 1384 EGF-LIKE 35.
FT DOMAIN 1387 1426 EGF-LIKE 36.
FT DOMAIN 1449 1462 CYS-RICH.
FT DOMAIN 1865 2076 6 X ANK MOTIF REPEATS.
FT REPEAT 1865 1910 ANK MOTIF 1.
FT REPEAT 1912 1942 ANK MOTIF 2.
FT REPEAT 1944 1975 ANK MOTIF 3.
FT REPEAT 1978 2009 ANK MOTIF 4.
FT REPEAT 2011 2042 ANK MOTIF 5.
FT REPEAT 2044 2076 ANK MOTIF 6.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.

Qy 1473 VSRLECRGGCAGGQC-----CGPLRSKRRKYSEFCTDG--SSFVDEV-----EKV 1515
| | : | | | |||| :: | : |
Db 890 SYRPLCAACYGCPRCPSDIDPCRPNCYN--CGSTTDYNAHAPDLDFCEACREPT 946

Qy 1516 VKC-----GCTRCV 1524
:| |
Db 947 NECATNPQNGANCTDCV 964

RESULT 5
NTC1_MOUSE
ID NTC1_MOUSE STANDARD; PRT; 2531 AA.
AC Q01705;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
GN NOTCH1 OR MOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
SEQUENCE FROM N.A.
RC TISSUE-EMBRYO;
RX MEDLINE-93194170; PubMed-8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.";
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 1551-2170 FROM N.A.
RC TISSUE-EMBRYO;
RX MEDLINE-93048835; PubMed-1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.";
RL Development 115:737-744(1992).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z11886; CAA77941.1; -.
DR HSSP; P00740; 1IXA.
DR MGD; MGI:97363; NOTCH1.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 35.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PRINTS; PRO0010; EGFBL00D.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.

FT	SIGNAL	1	18	POTENTIAL.
FT	CHAIN	19	2531	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT	DOMAIN	19	1725	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1726	1746	POTENTIAL.
FT	DOMAIN	1747	2531	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	1426	36 X EGF-TYPE REPEATS.
FT	DOMAIN	20	58	EGF-LIKE 1.
FT	DOMAIN	59	99	EGF-LIKE 2.
FT	DOMAIN	102	139	EGF-LIKE 3.
FT	DOMAIN	140	176	EGF-LIKE 4.
FT	DOMAIN	178	216	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	218	255	EGF-LIKE 6.
FT	DOMAIN	257	293	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	295	333	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	335	371	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372	410	EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	412	450	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	452	488	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	490	526	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	528	564	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	566	601	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	603	639	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	641	676	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	678	714	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	716	751	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	753	789	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	791	827	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	829	867	EGF-LIKE 22.
FT	DOMAIN	869	905	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	907	943	EGF-LIKE 24.
FT	DOMAIN	945	981	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	983	1019	EGF-LIKE 26.
FT	DOMAIN	1021	1057	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1059	1095	EGF-LIKE 28.
FT	DOMAIN	1097	1143	EGF-LIKE 29.
FT	DOMAIN	1145	1181	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1183	1219	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1221	1265	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1267	1305	EGF-LIKE 33.
FT	DOMAIN	1307	1346	EGF-LIKE 34.
FT	DOMAIN	1348	1384	EGF-LIKE 35.
FT	DOMAIN	1387	1426	EGF-LIKE 36.
FT	DOMAIN	1449	1462	CYS-RICH.
FT	DOMAIN	1445	1562	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1445	1480	LIN/NOTCH 1.
FT	REPEAT	1481	1522	LIN/NOTCH 2.
FT	REPEAT	1523	1562	LIN/NOTCH 3.
FT	DOMAIN	1865	2075	6 X ANK MOTIF REPEATS.
FT	REPEAT	1865	1910	ANK MOTIF 1.
FT	REPEAT	1912	1942	ANK MOTIF 2.
FT	REPEAT	1944	1975	ANK MOTIF 3.
FT	REPEAT	1978	2009	ANK MOTIF 4.
FT	REPEAT	2011	2042	ANK MOTIF 5.
FT	REPEAT	2044	2075	ANK MOTIF 6.
FT	DISULFID	24	37	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.
FT	DISULFID	245	254	BY SIMILARITY.
FT	DISULFID	261	272	BY SIMILARITY.
FT	DISULFID	266	281	BY SIMILARITY.

FT	DISULFID	283	292	BY SIMILARITY.
FT	DISULFID	299	312	BY SIMILARITY.
FT	DISULFID	306	321	BY SIMILARITY.
FT	DISULFID	323	332	BY SIMILARITY.
FT	DISULFID	339	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	361	370	BY SIMILARITY.
FT	DISULFID	376	387	BY SIMILARITY.
FT	DISULFID	381	398	BY SIMILARITY.
FT	DISULFID	400	409	BY SIMILARITY.
FT	DISULFID	416	429	BY SIMILARITY.
FT	DISULFID	423	438	BY SIMILARITY.
FT	DISULFID	440	449	BY SIMILARITY.
FT	DISULFID	456	467	BY SIMILARITY.
FT	DISULFID	461	476	BY SIMILARITY.
FT	DISULFID	478	487	BY SIMILARITY.
FT	DISULFID	494	505	BY SIMILARITY.
FT	DISULFID	499	514	BY SIMILARITY.
FT	DISULFID	516	525	BY SIMILARITY.
FT	DISULFID	532	543	BY SIMILARITY.
FT	DISULFID	537	552	BY SIMILARITY.
FT	DISULFID	554	563	BY SIMILARITY.
FT	DISULFID	570	580	BY SIMILARITY.
FT	DISULFID	575	589	BY SIMILARITY.
FT	DISULFID	591	600	BY SIMILARITY.
FT	DISULFID	607	618	BY SIMILARITY.
FT	DISULFID	612	627	BY SIMILARITY.
FT	DISULFID	629	638	BY SIMILARITY.
FT	DISULFID	645	655	BY SIMILARITY.
FT	DISULFID	650	664	BY SIMILARITY.
FT	DISULFID	666	675	BY SIMILARITY.
FT	DISULFID	682	693	BY SIMILARITY.
FT	DISULFID	687	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.
FT	DISULFID	720	730	BY SIMILARITY.
FT	DISULFID	725	739	BY SIMILARITY.
FT	DISULFID	741	750	BY SIMILARITY.
FT	DISULFID	757	768	BY SIMILARITY.
FT	DISULFID	762	777	BY SIMILARITY.
FT	DISULFID	779	788	BY SIMILARITY.
FT	DISULFID	795	806	BY SIMILARITY.
FT	DISULFID	800	815	BY SIMILARITY.
FT	DISULFID	817	826	BY SIMILARITY.
FT	DISULFID	833	844	BY SIMILARITY.
FT	DISULFID	838	855	BY SIMILARITY.
FT	DISULFID	857	866	BY SIMILARITY.
FT	DISULFID	873	884	BY SIMILARITY.
FT	DISULFID	878	893	BY SIMILARITY.
FT	DISULFID	895	904	BY SIMILARITY.
FT	DISULFID	911	922	BY SIMILARITY.
FT	DISULFID	916	931	BY SIMILARITY.
FT	DISULFID	933	942	BY SIMILARITY.
FT	DISULFID	987	998	BY SIMILARITY.
FT	DISULFID	992	1007	BY SIMILARITY.
FT	DISULFID	1009	1018	BY SIMILARITY.
FT	DISULFID	1025	1036	BY SIMILARITY.
FT	DISULFID	1030	1045	BY SIMILARITY.
FT	DISULFID	1047	1056	BY SIMILARITY.
FT	DISULFID	1063	1074	BY SIMILARITY.
FT	DISULFID	1068	1083	BY SIMILARITY.
FT	DISULFID	1085	1094	BY SIMILARITY.
FT	DISULFID	1101	1122	BY SIMILARITY.

Query Match 9.0%; Score 745; DB 1; Length 2531;
Best Local Similarity 24.4%; Pred. No. 1.3e-38;
Matches 238; Conservative 103; Mismatches 309; Indels 326; Gaps 46;

Qy 662 FNCNCYLAWLGWELRRKKRIVTGPNPCRKPYFLKEIPIQDVAIQDFTCDGDNDNSCSPLS 721
: | : | : | : | : | : | : | : | : | : |
Db 202 YRCACCAATHG-----PHCELPY-----VPCSPPSCNQNG---ATCRPTG 237

Qy 722 RCPTECTCLDTVV-----RCSNKGKLVLPKGI-----PRDVTLEYL--D 758
: | : | : | : | : | : | : | : | : | : |

[illegible]

```

RESULT 6
NTC1_HUMAN
ID NTC1_HUMAN STANDARD; PRT; 2444 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-
DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN NOTCH1 OR TAN1.

```


OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347367; PubMed=1831692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M73980; AAA60614.1; -.
DR HSSP; P00740; IIXA.
DR MIM; 190198; -.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 20.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 >2444 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT DOMAIN 19 1736 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1737 1757 POTENTIAL.
FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 EGF-LIKE 1.
FT DOMAIN 59 99 EGF-LIKE 2.
FT DOMAIN 102 139 EGF-LIKE 3.
FT DOMAIN 140 176 EGF-LIKE 4.
FT DOMAIN 178 216 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 218 255 EGF-LIKE 6.
FT DOMAIN 257 293 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 295 333 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 335 371 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 410 EGF-LIKE 10.
FT DOMAIN 412 450 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 452 488 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 490 526 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 528 564 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 566 601 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 603 639 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 641 676 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 678 714 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN	716	751	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	753	789	EGF-LIKE 20.
FT DOMAIN	791	827	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	829	868	EGF-LIKE 22.
FT DOMAIN	870	906	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	908	944	EGF-LIKE 24.
FT DOMAIN	946	982	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	984	1020	EGF-LIKE 26.
FT DOMAIN	1022	1058	EGF-LIKE 27.
FT DOMAIN	1060	1096	EGF-LIKE 28.
FT DOMAIN	1098	1144	EGF-LIKE 29.
FT DOMAIN	1146	1182	EGF-LIKE 30.
FT DOMAIN	1184	1220	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1222	1266	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	1268	1306	EGF-LIKE 33.
FT DOMAIN	1308	1347	EGF-LIKE 34.
FT DOMAIN	1349	1385	EGF-LIKE 35.
FT DOMAIN	1388	1427	EGF-LIKE 36.
FT DOMAIN	1446	1563	3 X LIN/NOTCH REPEATS.
FT REPEAT	1446	1481	LIN/NOTCH 1.
FT REPEAT	1482	1523	LIN/NOTCH 2.
FT REPEAT	1524	1563	LIN/NOTCH 3.
FT DOMAIN	1876	2087	6 X ANK MOTIF REPEATS.
FT REPEAT	1876	1921	ANK MOTIF 1.
FT REPEAT	1923	1954	ANK MOTIF 2.
FT REPEAT	1956	1987	ANK MOTIF 3.
FT REPEAT	1990	2021	ANK MOTIF 4.
FT REPEAT	2023	2054	ANK MOTIF 5.
FT REPEAT	2056	2087	ANK MOTIF 6.
FT DOMAIN	1576	1579	POLY-VAL.
FT DOMAIN	1662	1665	POLY-ARG.
FT DOMAIN	1729	1732	POLY-PRO.
FT DOMAIN	1741	1744	POLY-ALA.
FT DOMAIN	1902	1905	POLY-GLU.
FT DOMAIN	2260	2263	POLY-GLY.
FT DOMAIN	2404	2407	POLY-GLN.
FT DOMAIN	2411	2418	POLY-PRO.
FT DISULFID	24	37	BY SIMILARITY.
FT DISULFID	31	46	BY SIMILARITY.
FT DISULFID	48	57	BY SIMILARITY.
FT DISULFID	63	74	BY SIMILARITY.
FT DISULFID	68	87	BY SIMILARITY.
FT DISULFID	89	98	BY SIMILARITY.
FT DISULFID	106	117	BY SIMILARITY.
FT DISULFID	111	127	BY SIMILARITY.
FT DISULFID	129	138	BY SIMILARITY.
FT DISULFID	144	155	BY SIMILARITY.
FT DISULFID	149	164	BY SIMILARITY.
FT DISULFID	166	175	BY SIMILARITY.
FT DISULFID	182	195	BY SIMILARITY.
FT DISULFID	189	204	BY SIMILARITY.
FT DISULFID	206	215	BY SIMILARITY.
FT DISULFID	222	233	BY SIMILARITY.
FT DISULFID	227	243	BY SIMILARITY.
FT DISULFID	245	254	BY SIMILARITY.
FT DISULFID	261	272	BY SIMILARITY.
FT DISULFID	266	281	BY SIMILARITY.
FT DISULFID	283	292	BY SIMILARITY.
FT DISULFID	299	312	BY SIMILARITY.
FT DISULFID	306	321	BY SIMILARITY.
FT DISULFID	323	332	BY SIMILARITY.
FT DISULFID	339	350	BY SIMILARITY.
FT DISULFID	344	359	BY SIMILARITY.
FT DISULFID	361	370	BY SIMILARITY.
FT DISULFID	376	387	BY SIMILARITY.
FT DISULFID	381	398	BY SIMILARITY.
FT DISULFID	400	409	BY SIMILARITY.
FT DISULFID	416	429	BY SIMILARITY.
FT DISULFID	423	438	BY SIMILARITY.
FT DISULFID	440	449	BY SIMILARITY.
FT DISULFID	456	467	BY SIMILARITY.
FT DISULFID	461	476	BY SIMILARITY.
FT DISULFID	478	487	BY SIMILARITY.

```

662 FNCNCYLAWLGEWLRRKRVITGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDNNSCSPLS 721
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 YRCVC-----RATHGTGNPCERPY-----VPCSPSPQCNG---GTCRPTG 237

Qy 722 RCPTECTCLDVTV-----RCSNKGKLVLPKGIPIRDVTLEYLDGNQFTLVPK 767
      || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 DVTHECACLPGFTGQNCENIDDCPGNCKNGGACV-----DG----- 275

Qy 768 ELSNYKHLTLDLSNNRISTLSNQSFSNMQTLLTLILSYNRLCIPPRTFDGLKSLRLLS 827
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 276 -----VNTYN-----CPCPPWTG-----QYCT 293

Qy 828 LHGNDISVYPEGAFNDLSALSHLAIGANPLYDCDNQWLSDWVKSEYKEPGIARCAGPGE 887
      :: :: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 EDVDECQLMPNAQN---GGTCNTHGG---YNCVCVNGWTGEDCSENIDDCASAACFHGAT 349

Qy 888 MADKLLLTTPSKKFTCGQSPVD-VNILAKCN-PCLSNPCKNDGTCNSDPDFYR-CTCPYG 944
      || : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db 350 CHDRV-----ASYFCECPHGRTGLLCHLNDAISNPCKNEGSNCDTPVNGKAICTCPSG 403

Qy 945 FKGQDCDVPIHACI---SNPKCHGGTCHLKEGEEDGFWCICADGEGENCEFVNVDCCEDND 1002
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      404 YTGACSQVDVDECSLGANPCEHAGKCINTIG---SFECQCLQYGTGPCEIDVNECVSNP 460

Qy      1003 CENNSTCVDGINNYTCLCPPEYTGLCEEKLFCAQLDNPQCQHDSCILTPKGFKCDCTP 1062
       |:|::||:| | :|:| || || || || || || || || || || || ||
Db      461 CQNDAATCLDIGEFGCMCPGYEGVHCEVNTDECAS--SPLCHNGRCLDKINEFQCECPT 518

Qy      1063 GYVGHECHDIFDDCCDNKCKNGAHCTDAVNGYTICPEGYSGLFCESFPMPVLPRISPCD 1122
       |:| | | | :|:| ||||| | | |||:| |||:| || -- ||
Db      519 GFTGHLCKYDVDECASTPCKNKAACLDPNTYTVCTEGYTGHCEVDIDECDP--DPCH 576

Qy      1123 NFDCQGAQCIVRINEPICQCLPGYQGEKCEKLVSVNFINK-----SYLQI 1169
       |:| | | | | | ||| | || ||:| :| :| :| :| :|
Db      577 YGSKCDGVATFT-----CLCRPGYTGHCCE--TNINECSSQPCRLRGTCQDPDNAYL-- 626

Qy      1170 PSAKVRPQTNIQLTIADTE-----DSGILLYKGDKDHIAVELYGRVRASYDTGSHPAS 1223
       ::| | :| :| ||| | || | | | | | | | | | |
Db      627 -CFCLKGTGPNCEINLDDCASSPCDSGTLCDLKIDGYECACE-----PGY-TGSMCNS 677

Qy      1224 AI-----YSVETINDG-----NFHIVELLALDQSLSLVSDGGNPKIITNLKS 1265
       | | :| :| || :| :| :| :| :| :| :| :| :| :|
Db      678 NIDECAGNPCNHGGTGEDGINGFTCRCPEGYH-----DP--TCLSE 716

Qy      1266 QSTLNFDSPLYVGGMPGKNVASLRAPQGNGTSFHGCI RNLYINSELQDFQKVPMTGI 1325
       :| | :| :| | | || || | || :| :| :| :| :|
Db      717 VNECN-SNPVHGACRDSLNGYKCDPCPGWSGT-----NCDI NN----- 754

Qy      1326 LPGCEPCCHKVKCAH-GTCQPSSQAQGTCECGEGMWGPLCDQRNTDPLGNKCVH-GTCLP 1383
       | | :| :| ||:| :| | | ||:| || | | :| :| :| :| :|
Db      755 ---NECESNPNVNGGTCKDMT- SGIVCTRBFSGSPNCQTNINE-CASNPLNKGTCTID 808

Qy      1384 INAFSYSCKLEHGHHGLCDEEDLFNPPCAIKCKH-GKRCLSGLGOPY-CEC-SSGYTG 1440
       :| | || | | | | | | | | | | | | | | | | | |
Db      809 -DVAGYKNCNLLPYTGATC---EVVLAPFAPSPCRNGGECRSQSEDYESFSVCVPTAGAK 864

Qy      1441 DSCDREIS-----CRGERIRDYYQKQQGYAACOTTKKVSRLECRGGCAGGC-----C 1488
       :|:| :| || || | | | | | | | | | | | | | |
Db      865 QTCEVDINECVLSPCR-----HG-ASCQNTHGXYRCHCQAGYSGRNCE TIDDC 912

Qy      1489 GPLRSKRRKYSFECTDG---SSFVDEV-----EKVVKC-----GCTRCV 1524
       |||| | | :| :| :| :| :| :| :| :| :| :|
Db      913 RPNPCHN---GGSCDTGINTAFCDCLPGFRGTGCEEDINECASDPCRNGANCPTCY 965

```

CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURA

CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC BRAIN AND HEAD REGIONS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X69088; CAA48831.1; -.
DR HSSP; P00740; IIXA.
DR ZFIN; ZDB-GENE-990415-173; NOTCH.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001336; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 36.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 3.
DR PRINTS; PR00009; EGFTGF.
DR PRINTS; PR00010; EGFBLD.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS01187; EGF_CA; 22.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1725 1747 POTENTIAL.
FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 57 EGF-LIKE 1.
FT DOMAIN 58 98 EGF-LIKE 2.
FT DOMAIN 101 138 EGF-LIKE 3.
FT DOMAIN 139 175 EGF-LIKE 4.
FT DOMAIN 177 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 294 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 489 524 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 526 562 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 564 599 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 601 637 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 639 674 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 676 712 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 714 749 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 751 787 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 827 865 EGF-LIKE 22.
FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 981 1017 EGF-LIKE 26.
FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1057 1093 EGF-LIKE 28.
FT DOMAIN 1095 1141 EGF-LIKE 29.
FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1265 1303 EGF-LIKE 33.
FT DOMAIN 1305 1344 EGF-LIKE 34.
FT DOMAIN 1346 1382 EGF-LIKE 35.
FT DOMAIN 1385 1423 EGF-LIKE 36.
FT DOMAIN 1446 1561 3 X LIN/NOTCH REPEATS.
FT REPEAT 1446 1486 LIN/NOTCH 1.
FT REPEAT 1487 1520 LIN/NOTCH 2.
FT REPEAT 1521 1561 LIN/NOTCH 3.
FT DOMAIN 1861 2074 6 X ANK MOTIF REPEATS.
FT REPEAT 1861 1891 ANK MOTIF 1.
FT REPEAT 1892 1940 ANK MOTIF 1.
FT REPEAT 1941 1974 ANK MOTIF 1.
FT REPEAT 1975 2007 ANK MOTIF 1.
FT REPEAT 2008 2040 ANK MOTIF 1.
FT REPEAT 2041 2074 ANK MOTIF 1.
FT DOMAIN 2265 2276 POLY-GLN (OPA-REPEAT).
FT DISULFID 25 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 86 BY SIMILARITY.
FT DISULFID 88 97 BY SIMILARITY.
FT DISULFID 105 116 BY SIMILARITY.
FT DISULFID 110 126 BY SIMILARITY.
FT DISULFID 128 137 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 181 194 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 578 BY SIMILARITY.
FT DISULFID 573 587 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the uEGF gene in the sea urchin
RT stronglyloccentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.";
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.";
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin.";
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.";
RL Dev. Biol. 146:89-99(1991).
CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; IA (SHOWN HERE) AND IB; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF
CC REPEATS.
CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L08692; AAA62164.1; -.
DR EMBL; L08692; AAA62163.1; -.
DR EMBL; X17530; CAA35571.1; -.
DR EMBL; M17421; AAA30050.1; -.
DR EMBL; X17533; CAA35573.1; -.
DR PIR; A29316; A29316.
DR HSPSP; P01132; IEPH.
DR INTERPRO; IPR000088; -.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000859; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF01382; Avidin; 1.
DR PFAM; PF00431; CUB; 1.
DR PFAM; PF00008; EGF; 21.

DR PRINTS; PR00010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; 19.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 19.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
KW Glycoprotein.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1064 FIBROPELLIN I.
FT DOMAIN 20 55 EGF-LIKE 1.
FT DOMAIN 62 175 CUB.
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 860 896 EGF-LIKE 20.
FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 936 1064 AVIDIN-LIKE.
FT DISULFID 23 34 BY SIMILARITY.
FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 240 249 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 287 BY SIMILARITY.
FT DISULFID 294 305 BY SIMILARITY.
FT DISULFID 299 314 BY SIMILARITY.
FT DISULFID 316 325 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.

```

RESULT 9
NTC3_MOUSE
ID NTC3_MOUSE STANDARD; PRT; 2318 AA.
AC Q61982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEUROGENIC LOCUS NOTCH 3 PROTEIN.
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR X SWISS WEBSTER;
RX MEDLINE=95001556; PubMed=7918097;
RA Lardelli M., Dalstrand J., Lendahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium.";
RL Mech. Dev. 46:123-136(1994).
CC -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74760; CAA52776.1; -.
DR HSSP; P00740; LIXA.
DR MGD; MGI:99460; NOTCH3.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 34.
DR PFAM; PF00023;  $\alpha$ nk; 6.
DR PFAM; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 17.
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein.
FT DOMAIN 1 1643 EXTRACELLULAR.
FT TRANSMEM 1644 1664 POTENTIAL.

```

FT	DOMAIN	1665	2318	CYTOPLASMIC.
FT	DOMAIN	39	1374	34 X EGF-TYPE REPEATS.
FT	DOMAIN	1388	1503	3 X LIN/NOTCH REPEATS.
FT	DOMAIN	1784	1998	6 X CDC10/SWI6 REPEATS.
FT	DOMAIN	2242	2261	PEST.
FT	DOMAIN	39	78	EGF-LIKE 1.
FT	DOMAIN	79	119	EGF-LIKE 2.
FT	DOMAIN	120	157	EGF-LIKE 3.
FT	DOMAIN	159	196	EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	198	235	EGF-LIKE 5.
FT	DOMAIN	237	273	EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	275	313	EGF-LIKE 7.
FT	DOMAIN	315	351	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	390	EGF-LIKE 9.
FT	DOMAIN	392	430	EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	432	468	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	470	506	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	508	544	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	546	581	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	583	619	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	621	656	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	658	694	EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	696	731	EGF-LIKE 18.
FT	DOMAIN	735	771	EGF-LIKE 19.
FT	DOMAIN	772	809	EGF-LIKE 20.
FT	DOMAIN	811	848	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	850	886	EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	888	923	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	925	961	EGF-LIKE 24.
FT	DOMAIN	963	999	EGF-LIKE 25.
FT	DOMAIN	1001	1035	EGF-LIKE 26.
FT	DOMAIN	1037	1083	EGF-LIKE 27.
FT	DOMAIN	1085	1121	EGF-LIKE 28.
FT	DOMAIN	1123	1159	EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1161	1204	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1206	1245	EGF-LIKE 31.
FT	DOMAIN	1247	1288	EGF-LIKE 32.
FT	DOMAIN	1290	1326	EGF-LIKE 33.
FT	DOMAIN	1336	1374	EGF-LIKE 34.
FT	REPEAT	1388	1428	LIN/NOTCH 1.
FT	REPEAT	1429	1467	LIN/NOTCH 2.
FT	REPEAT	1468	1503	LIN/NOTCH 3.
FT	REPEAT	1784	1816	CDC10/SWI6 1.
FT	REPEAT	1817	1865	CDC10/SWI6 2.
FT	REPEAT	1866	1898	CDC10/SWI6 3.
FT	REPEAT	1899	1932	CDC10/SWI6 4.
FT	REPEAT	1933	1965	CDC10/SWI6 5.
FT	REPEAT	1966	1998	CDC10/SWI6 6.
FT	DISULFID	43	55	BY SIMILARITY.
FT	DISULFID	49	66	BY SIMILARITY.
FT	DISULFID	68	77	BY SIMILARITY.
FT	DISULFID	83	94	BY SIMILARITY.
FT	DISULFID	88	107	BY SIMILARITY.
FT	DISULFID	109	118	BY SIMILARITY.
FT	DISULFID	124	135	BY SIMILARITY.
FT	DISULFID	129	145	BY SIMILARITY.
FT	DISULFID	147	156	BY SIMILARITY.
FT	DISULFID	163	175	BY SIMILARITY.
FT	DISULFID	169	184	BY SIMILARITY.
FT	DISULFID	186	195	BY SIMILARITY.
FT	DISULFID	202	213	BY SIMILARITY.
FT	DISULFID	207	223	BY SIMILARITY.
FT	DISULFID	225	234	BY SIMILARITY.
FT	DISULFID	241	252	BY SIMILARITY.
FT	DISULFID	246	261	BY SIMILARITY.
FT	DISULFID	263	272	BY SIMILARITY.
FT	DISULFID	279	292	BY SIMILARITY.
FT	DISULFID	286	301	BY SIMILARITY.
FT	DISULFID	303	312	BY SIMILARITY.
FT	DISULFID	319	330	BY SIMILARITY.
FT	DISULFID	324	339	BY SIMILARITY.
FT	DISULFID	341	350	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.

FT	DISULFID	361	378	BY SIMILARITY.
FT	DISULFID	380	389	BY SIMILARITY.
FT	DISULFID	396	409	BY SIMILARITY.
FT	DISULFID	403	418	BY SIMILARITY.
FT	DISULFID	420	429	BY SIMILARITY.
FT	DISULFID	436	447	BY SIMILARITY.
FT	DISULFID	441	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	474	485	BY SIMILARITY.
FT	DISULFID	479	494	BY SIMILARITY.
FT	DISULFID	496	505	BY SIMILARITY.
FT	DISULFID	512	523	BY SIMILARITY.
FT	DISULFID	517	532	BY SIMILARITY.
FT	DISULFID	534	543	BY SIMILARITY.
FT	DISULFID	550	560	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	571	580	BY SIMILARITY.
FT	DISULFID	587	598	BY SIMILARITY.
FT	DISULFID	592	607	BY SIMILARITY.
FT	DISULFID	609	618	BY SIMILARITY.
FT	DISULFID	625	635	BY SIMILARITY.
FT	DISULFID	630	644	BY SIMILARITY.
FT	DISULFID	646	655	BY SIMILARITY.
FT	DISULFID	662	673	BY SIMILARITY.
FT	DISULFID	667	682	BY SIMILARITY.
FT	DISULFID	684	693	BY SIMILARITY.
FT	DISULFID	700	710	BY SIMILARITY.
FT	DISULFID	705	719	BY SIMILARITY.
FT	DISULFID	721	730	BY SIMILARITY.
FT	DISULFID	739	750	BY SIMILARITY.
FT	DISULFID	744	759	BY SIMILARITY.
FT	DISULFID	761	770	BY SIMILARITY.
FT	DISULFID	776	787	BY SIMILARITY.
FT	DISULFID	781	797	BY SIMILARITY.
FT	DISULFID	799	808	BY SIMILARITY.
FT	DISULFID	815	827	BY SIMILARITY.
FT	DISULFID	821	836	BY SIMILARITY.
FT	DISULFID	838	847	BY SIMILARITY.
FT	DISULFID	854	865	BY SIMILARITY.
FT	DISULFID	859	874	BY SIMILARITY.
FT	DISULFID	876	885	BY SIMILARITY.
FT	DISULFID	892	902	BY SIMILARITY.
FT	DISULFID	897	911	BY SIMILARITY.
FT	DISULFID	913	922	BY SIMILARITY.
FT	DISULFID	929	940	BY SIMILARITY.
FT	DISULFID	934	949	BY SIMILARITY.
FT	DISULFID	951	960	BY SIMILARITY.
FT	DISULFID	967	978	BY SIMILARITY.
FT	DISULFID	972	987	BY SIMILARITY.
FT	DISULFID	989	998	BY SIMILARITY.
FT	DISULFID	1005	1016	BY SIMILARITY.
FT	DISULFID	1010	1023	BY SIMILARITY.
FT	DISULFID	1025	1034	BY SIMILARITY.
FT	DISULFID	1041	1062	BY SIMILARITY.
FT	DISULFID	1056	1071	BY SIMILARITY.
FT	DISULFID	1073	1082	BY SIMILARITY.
FT	DISULFID	1089	1100	BY SIMILARITY.
FT	DISULFID	1094	1109	BY SIMILARITY.
FT	DISULFID	1111	1120	BY SIMILARITY.
FT	DISULFID	1127	1138	BY SIMILARITY.
FT	DISULFID	1132	1147	BY SIMILARITY.
FT	DISULFID	1149	1158	BY SIMILARITY.
FT	DISULFID	1165	1183	BY SIMILARITY.
FT	DISULFID	1177	1192	BY SIMILARITY.
FT	DISULFID	1194	1203	BY SIMILARITY.
FT	DISULFID	1210	1223	BY SIMILARITY.
FT	DISULFID	1215	1233	BY SIMILARITY.
FT	DISULFID	1235	1244	BY SIMILARITY.

Query Match 8.7%; Score 722; DB 1; Length 2318;
 Best Local Similarity 23.7%; Pred. No. 3.1e-37;
 Matches 227; Conservative 86; Mismatches 288; Indels 358; Gaps 37;

```

RESULT 10
NTC4_MOUSE
ID   NTC4_MOUSE          STANDARD;          PRT;   1964 AA.
AC   P31695; 062389;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4 PRECURSOR (TRANSFORMING

```

```

DE PROTEIN INT-3).
GN NOTCH4 OR INT3 OR INT-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92194507; PubMed=1312643;
RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT "Mouse mammary tumor gene int-3: a member of the notch gene family
RT transforms mammary epithelial cells.";
RL J. Virol. 66:2594-2599(1992).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=97294599; PubMed=9150355;
RA Gallahan D., Callahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
RT the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1883-1890(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG, AND TESTIS;
RX MEDLINE=96281668; PubMed=8681805;
RA Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene.";
RL Development 122:2251-2259(1996).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DISEASE: ACTIVATED INT-3 TRANSFORMS MAMMARY EPITHELIAL CELLS.
CC -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 CDC10/SWI6 REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80456; AAB38377.1; -.
DR EMBL; U43691; AAC52630.1; -.
DR PIR; A38072; TVMVT3.
DR HSSP; P00740; 1IXA.
DR MGD; MGI:107471; NOTCH4.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000800; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002110; -.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00023; ank; 6.
DR PFAM; PF00066; notch; 2.
DR PRINTS; PR00010; EGFkBLOOD.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 11.
DR PROSITE; PS00022; EGF_1; 28.
DR PROSITE; PS01186; EGF_2; 21.
DR PROSITE; PS01187; EGF_CA; 9.
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Proto-oncogene; ANK repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT DOMAIN 21 1443 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1444 1464 POTENTIAL.
FT DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 60 EGF-LIKE 1.
FT DOMAIN 61 112 EGF-LIKE 2.
FT DOMAIN 115 152 EGF-LIKE 3.

```


FT	DOMAIN	153	189	EGF-LIKE 4.
FT	DOMAIN	191	229	EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	231	271	EGF-LIKE 6.
FT	DOMAIN	273	309	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	311	350	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	352	388	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	389	427	EGF-LIKE 10.
FT	DOMAIN	429	470	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	472	508	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	510	546	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	548	584	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	586	622	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	622	656	EGF-LIKE 16.
FT	DOMAIN	658	686	EGF-LIKE 17.
FT	DOMAIN	688	724	EGF-LIKE 18.
FT	DOMAIN	726	762	EGF-LIKE 19.
FT	DOMAIN	764	800	EGF-LIKE 20.
FT	DOMAIN	803	839	EGF-LIKE 21.
FT	DOMAIN	841	877	EGF-LIKE 22.
FT	DOMAIN	878	924	EGF-LIKE 23.
FT	DOMAIN	926	962	EGF-LIKE 24.
FT	DOMAIN	964	1000	EGF-LIKE 25.
FT	DOMAIN	1002	1040	EGF-LIKE 26.
FT	DOMAIN	1042	1081	EGF-LIKE 27.
FT	DOMAIN	1083	1122	EGF-LIKE 28.
FT	DOMAIN	1126	1167	EGF-LIKE 29.
FT	DOMAIN	1168	1282	3 X LIN/NOTCH REPEATS.
FT	REPEAT	1168	1208	LIN/NOTCH 1.
FT	REPEAT	1209	1242	LIN/NOTCH 2.
FT	REPEAT	1243	1282	LIN/NOTCH 3.
FT	DOMAIN	1572	1785	6 X ANK MOTIF REPEATS.
FT	REPEAT	1572	1603	ANK MOTIF 1.
FT	REPEAT	1622	1653	ANK MOTIF 2.
FT	REPEAT	1654	1685	ANK MOTIF 3.
FT	REPEAT	1688	1719	ANK MOTIF 4.
FT	REPEAT	1721	1752	ANK MOTIF 5.
FT	REPEAT	1754	1785	ANK MOTIF 6.
FT	DISULFID	25	38	BY SIMILARITY.
FT	DISULFID	32	48	BY SIMILARITY.
FT	DISULFID	50	59	BY SIMILARITY.
FT	DISULFID	65	77	BY SIMILARITY.
FT	DISULFID	71	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	130	BY SIMILARITY.
FT	DISULFID	124	140	BY SIMILARITY.
FT	DISULFID	142	151	BY SIMILARITY.
FT	DISULFID	157	168	BY SIMILARITY.
FT	DISULFID	162	177	BY SIMILARITY.
FT	DISULFID	179	188	BY SIMILARITY.
FT	DISULFID	195	208	BY SIMILARITY.
FT	DISULFID	202	217	BY SIMILARITY.
FT	DISULFID	219	228	BY SIMILARITY.
FT	DISULFID	235	246	BY SIMILARITY.
FT	DISULFID	240	259	BY SIMILARITY.
FT	DISULFID	261	270	BY SIMILARITY.
FT	DISULFID	277	288	BY SIMILARITY.
FT	DISULFID	282	297	BY SIMILARITY.
FT	DISULFID	299	308	BY SIMILARITY.
FT	DISULFID	315	329	BY SIMILARITY.
FT	DISULFID	323	338	BY SIMILARITY.
FT	DISULFID	340	349	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	376	BY SIMILARITY.
FT	DISULFID	378	387	BY SIMILARITY.
FT	DISULFID	393	404	BY SIMILARITY.
FT	DISULFID	398	415	BY SIMILARITY.
FT	DISULFID	417	426	BY SIMILARITY.
FT	DISULFID	433	449	BY SIMILARITY.
FT	DISULFID	443	458	BY SIMILARITY.
FT	DISULFID	460	469	BY SIMILARITY.
FT	DISULFID	476	487	BY SIMILARITY.
FT	DISULFID	481	496	BY SIMILARITY.
FT	DISULFID	498	507	BY SIMILARITY.

FT	DISULFID	514	525	BY SIMILARITY.
FT	DISULFID	519	534	BY SIMILARITY.
FT	DISULFID	536	545	BY SIMILARITY.
FT	DISULFID	552	563	BY SIMILARITY.
FT	DISULFID	557	572	BY SIMILARITY.
FT	DISULFID	574	583	BY SIMILARITY.
FT	DISULFID	590	601	BY SIMILARITY.
FT	DISULFID	595	610	BY SIMILARITY.
FT	DISULFID	612	621	BY SIMILARITY.
FT	DISULFID	626	637	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	648	655	BY SIMILARITY.
FT	DISULFID	662	669	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	692	703	BY SIMILARITY.
FT	DISULFID	697	712	BY SIMILARITY.
FT	DISULFID	714	723	BY SIMILARITY.
FT	DISULFID	730	741	BY SIMILARITY.
FT	DISULFID	735	750	BY SIMILARITY.
FT	DISULFID	752	761	BY SIMILARITY.
FT	DISULFID	768	779	BY SIMILARITY.
FT	DISULFID	773	788	BY SIMILARITY.
FT	DISULFID	790	799	BY SIMILARITY.
FT	DISULFID	807	818	BY SIMILARITY.
FT	DISULFID	812	827	BY SIMILARITY.
FT	DISULFID	829	838	BY SIMILARITY.
FT	DISULFID	845	856	BY SIMILARITY.
FT	DISULFID	850	865	BY SIMILARITY.
FT	DISULFID	867	876	BY SIMILARITY.
FT	DISULFID	882	903	BY SIMILARITY.
FT	DISULFID	897	912	BY SIMILARITY.
FT	DISULFID	914	923	BY SIMILARITY.
FT	DISULFID	930	941	BY SIMILARITY.
FT	DISULFID	935	950	BY SIMILARITY.
FT	DISULFID	952	961	BY SIMILARITY.
FT	DISULFID	968	979	BY SIMILARITY.
FT	DISULFID	973	988	BY SIMILARITY.
FT	DISULFID	990	999	BY SIMILARITY.
FT	DISULFID	1006	1019	BY SIMILARITY.
FT	DISULFID	1011	1028	BY SIMILARITY.
FT	DISULFID	1030	1039	BY SIMILARITY.
FT	DISULFID	1046	1057	BY SIMILARITY.
FT	DISULFID	1051	1069	BY SIMILARITY.
FT	DISULFID	1071	1080	BY SIMILARITY.
FT	DISULFID	1087	1098	BY SIMILARITY.
FT	DISULFID	1092	1110	BY SIMILARITY.
FT	DISULFID	1112	1121	BY SIMILARITY.

Query Match 8.3%; Score 691.5; DB 1; Length 1964;
Best Local Similarity 24.8%; Pred. No. 2e-35;
Matches 235; Conservative 86; Mismatches 345; Indels 283; Gaps 41;

Qy	664	CNCYLAWLGEWLAKKRIVTGNPRCQKPYFLKEIPQDVAIQDFTCDDGNDNDCS----	718
Db	48	CQCAPGFLGE-----TCQFP-----DPCRDQL-----CKNG---GSCQALLPT	83
Qy	719	-PLSRCPT-----ECTCLDTVVRCSNKGVLKVPKGIPTVTELYLDGNQFTLVPKELS	770
Db	84	PPSSRSPTSPLTPHESCTC-----PSGFTGDRQCRLHLEE---LCPPSFCS	125
Qy	771	NYKHLTLIDLNSNNRISTLSNQSFNSMTQL-----TLILSYNRLRCIPRPTFD	818
Db	126	NGGH-CYQASGRPQCSCEPGWTGEQCQLRDFCSANPCANGGVCLATYPQIQCRCPGFE	184
Qy	819	GLKSLRLSLAGNDISVVPEGAFNDLSALSHLAIGANPLYCDNCNQLSDWKSEYKEPG	878
Db	185	GHTCERDINECFLEPGPCPQG-----TSCHTNLGSYQCLCPVGQEGPQCKLRKGACPPG	238
Qy	879	IARCAG-----PGEMADKLLLTTPSKKFTCQGPVDVNILAKCNP--CLSNPCKNDGTCN	930
Db	239	SCLNGGTCLVPEGHSTFHLCLCPPG--FT-----GLDCMNPDCCVRHQCQNGATC-	288

QY 931 SDPVDFYRCTCPYGFKGQDCDVIHACISN---PCKHGGTCHLKEGEEDGFWICADGFE 987
Db 289 LDGLDTYTLCPKTKWKGDCSEDIDECARGPPRCRNGGTCQNTAG---SFHCVCVSWG 345
QY 988 GENCEVNVDDCEDNDCENNSTCVDGINNYTCLCP-----1021
Db 346 GAGCEENLDDCAATCAPGSTCIDRVGSFSLCPPGRTGLLCHLEDMLSQPCHVNAQS 405
QY 1022 -----PEYTGELCEEKLDPC---AQDLNQCQHDSKCILTPKGFKCDCTPGYV 1065
Db 406 TNPLTGTSLCICQPGYSGSTCHQDLDECQMAQQGSPSPCEHGGSCINTPGSFNCLCLPGYT 465
QY 1066 GEHCIDFDDCDQNKCKNGAHTDAVNGYTCICPEGYSGLFCEFPMPVLPRTSPCDNFD 1125
Db 466 GSRCEADHNECLSQPCHPGSTCLDLLATFHCLCPGLGRLCEV-----EVNCTSNP 518
QY 1126 CQGAQCIVRINEPICQCLPGYQGEKELVSVNFINKESYLQIPSAKVRPQTNITLQIA 1185
Db 519 CLNQACHDLLNGFQCLCLPGFTGARCEK-----D 548
QY 1186 TDEDSGILLYKGDKDHIAVELYRGRVRSYDTGSHPASAIYSVETINDGNFHVIVELL---1242
Db 549 MDECSSTPCANG-----GRCRD-----QPGAFY-CECLPGF 578
QY 1243 -----ALDQSL-----LSVDGG-----NPKIITNLSKQSTLNFDPLVYVGM- 1280
Db 579 EGPHECEVDECLSDPCPVGASCLDLPGAFFCLCRGFTGQLC-----EVLCTPNMC 631
QY 1281 -PGKSNVASLRQAP-----GQNGTSFHG-CIRNLYINSELQDFQKVPMTGI 1325
Db 632 QPGQQCQGEHRAPCLCPDGPSPGCPAEDNCPCHHGHCQRSLCVDE-----GWTG- 682
QY 1326 LPGCEP----CHKKVCAH-GTCQPSQAGFTCEQEGWGPLCDQRTNDPLGNKCVHGT 1380
Db 683 -PECETELGGCISTPCAAGGTCHP-QPSGYNCTCPAGYMGITCSEEV-ACHSGPLNGG 739
QY 1381 CLPINAFYSCKCLEGHGVLCDDEEDLFNCPQAIKCKHGKRLSGLQGPYCESSGYTG 1440
Db 740 SCSIRPEYSCTCLPSHTGRHCQTAVD---HCVSASCLNGGTCVNKPGTFFCLCATGQFG 796
QY 1441 DSCDREI--SCRGERIRDYKQKQGYAACQTTKKVSRLECRGGCAGGQC 1487
Db 797 LHCEERTNPSCADSPCRN-----KATCQDTPRGARCLCSPGYTGSSC 838

RESULT 11

CRB_DROME

ID CRB_DROME STANDARD; PRT; 2139 AA.
AC P10040;
DT 01-MAR-1989 (Rel. 10, Created)
01-MAY-1991 (Rel. 18, Last sequence update)
15-DEC-1998 (Rel. 37, Last annotation update)
DE CRUMBS PROTEIN PRECURSOR (95F).
GN CRB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON-R; TISSUE-EMBRYO;
RX MEDLINE=90263104; PubMed=2344615;
RA Tepass U., Theres C., Knust E.;
RT "Crums encodes an EGF-like protein expressed on apical membranes of
Drosophila epithelial cells and required for organization of
epithelia".
RL Cell 61:787-799(1990).
RN [2]
RP SEQUENCE OF 1663-1955 FROM N.A.
RX MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
Vaessin H., Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila

RT melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC POLARITY. IT MAY ACT AS A SIGNAL.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PTM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -.
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR HSSP; P00740; 1XA.
DR FLYBASE; FBgn0000368; crb.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00054; laminin_G; 3.
DR PRINTS; PR00010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
KW Differentiation; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 91 2139 CRUMBS PROTEIN.
FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.
FT DOMAIN 348 386 EGF-LIKE 3.
FT DOMAIN 388 425 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 427 463 EGF-LIKE 5.
FT DOMAIN 464 500 EGF-LIKE 6.
FT DOMAIN 501 532 EGF-LIKE 7.
FT DOMAIN 545 581 EGF-LIKE 8.
FT DOMAIN 582 611 EGF-LIKE 9.
FT DOMAIN 609 646 EGF-LIKE 10.
FT DOMAIN 648 685 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 687 723 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 725 761 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 763 800 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 802 838 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 840 902 EGF-LIKE 16.
FT DOMAIN 904 940 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 942 978 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 980 1021 EGF-LIKE 19.
FT DOMAIN 1207 1243 EGF-LIKE 20.
FT DOMAIN 1481 1517 EGF-LIKE 21.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1797 1833 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1835 1871 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1874 1915 EGF-LIKE 25.
FT DOMAIN 1915 1951 EGF-LIKE 26.
FT DOMAIN 1953 1989 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1991 2029 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2030 2070 EGF-LIKE 29.
FT DISULFID 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.

Qy 1121 C-----DNFDCQ-----NGAQCI-VRINEP----ICQC 1143
 Db 952 CQRGTCYDQDDYDCDCDANYGGKNCVLLKGCQNPCLNGGACLPYLINEVTHLYTCTC 1011

Qy 1144 LPGYQGEKCEK-----LVSFNFINKESYLQIPSAKVRPQTNITLQIATDEDSGIL 1193
 Db 1012 ENGFGQDKCEKTTLSMVATSLISVTTREEGY-----DINLQFRTTLPLNGVL 1059

Qy 1194 LY---KGDGDH---IAYLYRGRVRSYDTGSHPASAIYSVETINDGNFHVALLDQSL 1248
 Db 1060 AFGTTGKNEPVSYLELINGRLNL-HSSLLNKWGVFIGSKLNDNSNWHKV-FVAINTS- 1116

Qy 1249 SLSVDGNGPKIITNLKQSTLNFDSP----LYVGG-MPG-KSNVASLRQAPGQNGTSFHH 1302
 Db 1117 HVLVSANDEQAIPVGSYETANNSQSPFRTYLGCTIPNLKSYLRHLTHQP----SAFVG 1172

Qy 1303 CIRNLYINSE--LQDFQKVPQM-----TGILPGC-----EPCHKV-CA---HGTCQP 1344
 Db 1173 CMQDIMVNGKWIFFDEQDANISYTKLENVQSGCPRTEQCKPNPCHSNVECTDLNH----- 1227

Qy 1345 SSQAGFTCECEGWMGPLCDQRTNDPCLGNK 1375
 Db 1228 ----TFACHCPRPFPGHTCQHNMTAATFGHE 1254

RESULT 12

SERR_DROME

ID SERR_DROME STANDARD; PRT; 1408 AA.
 AC P18168;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SERRATE PROTEIN PRECURSOR (BEADED PROTEIN).
 GN SER OR BD.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE=91347903; PubMed=1840519;
 RA Thomas U., Speicher S.A., Knust E.;
 RT "The Drosophila gene Serrate encodes an EGF-like transmembrane
 protein with a complex expression pattern in embryos and wing
 discs.";
 RL Development 111:749-761(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91099666; PubMed=2125287;
 RA Fleming R.J., Scottgale T.N., Diederich R.J., Artavanis-Tsakonas S.;
 RT "The gene Serrate encodes a putative EGF-like transmembrane protein
 essential for proper ectodermal development in Drosophila
 melanogaster.";
 RL Genes Dev. 4:2188-2201(1990).
 CC -!- FUNCTION: ESSENTIAL FOR PROPER ECTODERMAL DEVELOPMENT. SERRATE
 CC MAY REPRESENT AN ELEMENT IN A NETWORK OF INTERACTING MOLECULES
 CC OPERATING AT THE CELL SURFACE DURING THE DIFFERENTIATION OF
 CC CERTAIN TISSUES.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: APPEARS TO BE RESTRICTED EXCLUSIVELY TO
 CC CELLS OF ECTODERMAL ORIGIN.
 CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.
 CC -!- MISCELLANEOUS: NOTCH AND SERRATE MAY INTERACT AT THE PROTEIN
 CC LEVEL, IT IS CONCEIVABLE THAT THE SERRATE AND DELTA PROTEINS MAY
 CC COMPETE FOR BINDING WITH THE NOTCH PROTEIN.
 CC -!- SIMILARITY: CONTAINS 14 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE DELTA/SERRATE/JAGGED FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X56811; CAA40148.1; -.
 DR EMBL; M35759; AAA28938.1; -.
 DR PIR; A36666; A36666.
 DR PIR; S16878; S16878.
 DR HSSP; P00743; 1WHE.
 DR FLYBASE; FBgn0004197; Ser.
 DR INTERPRO; IPR000152; -.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR001438; -.
 DR INTERPRO; IPR001774; -.
 DR INTERPRO; IPR001881; -.
 DR PFAM; PF01414; DSL; 1.
 DR PFAM; PF00008; EGF; 11.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PROSITE; PS00010; ASX_HYDROXYL; 7.
 DR PROSITE; PS00022; EGF_1; 14.
 DR PROSITE; PS01186; EGF_2; 8.
 DR PROSITE; PS01187; EGF_CA; 5.
 KW Differentiation; Repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 83 POTENTIAL.
 FT CHAIN 84 1408 SERRATE PROTEIN.
 FT DOMAIN 84 1223 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1224 1249 POTENTIAL.
 FT DOMAIN 1250 1408 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 284 317 EGF-LIKE 1.
 FT DOMAIN 315 349 EGF-LIKE 2.
 FT DOMAIN 351 389 EGF-LIKE 3.
 FT DOMAIN 391 489 EGF-LIKE 4.
 FT DOMAIN 407 476 SER-RICH (INSERT).
 FT DOMAIN 491 527 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 529 609 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 611 646 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 648 684 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 686 721 EGF-LIKE 9.
 FT DOMAIN 723 797 EGF-LIKE 10.
 FT DOMAIN 737 769 THR-RICH (INSERT).
 FT DOMAIN 799 835 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 837 877 EGF-LIKE 12.
 FT DOMAIN 879 915 EGF-LIKE 13.
 FT DOMAIN 917 953 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 288 299 BY SIMILARITY.
 FT DISULFID 292 305 BY SIMILARITY.
 FT DISULFID 307 316 BY SIMILARITY.
 FT DISULFID 319 330 BY SIMILARITY.
 FT DISULFID 325 337 BY SIMILARITY.
 FT DISULFID 339 348 BY SIMILARITY.
 FT DISULFID 355 367 BY SIMILARITY.
 FT DISULFID 361 377 BY SIMILARITY.
 FT DISULFID 379 388 BY SIMILARITY.
 FT DISULFID 395 406 BY SIMILARITY.
 FT DISULFID 400 477 BY SIMILARITY.
 FT DISULFID 479 488 BY SIMILARITY.
 FT DISULFID 495 506 BY SIMILARITY.
 FT DISULFID 500 515 BY SIMILARITY.
 FT DISULFID 517 526 BY SIMILARITY.
 FT DISULFID 533 588 BY SIMILARITY.
 FT DISULFID 582 597 BY SIMILARITY.
 FT DISULFID 599 608 BY SIMILARITY.
 FT DISULFID 615 625 BY SIMILARITY.
 FT DISULFID 619 634 BY SIMILARITY.
 FT DISULFID 636 645 BY SIMILARITY.
 FT DISULFID 652 663 BY SIMILARITY.
 FT DISULFID 657 672 BY SIMILARITY.
 FT DISULFID 674 683 BY SIMILARITY.
 FT DISULFID 690 700 BY SIMILARITY.

FT DISULFID 695 709 BY SIMILARITY.
FT DISULFID 711 720 BY SIMILARITY.
FT DISULFID 803 814 BY SIMILARITY.
FT DISULFID 808 823 BY SIMILARITY.
FT DISULFID 825 834 BY SIMILARITY.
FT DISULFID 841 852 BY SIMILARITY.
FT DISULFID 846 865 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 883 894 BY SIMILARITY.
FT DISULFID 888 903 BY SIMILARITY.
FT DISULFID 905 914 BY SIMILARITY.
FT DISULFID 921 932 BY SIMILARITY.
FT DISULFID 926 941 BY SIMILARITY.
FT DISULFID 943 952 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 247 247 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 739 739 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 977 977 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1004 1004 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 14 17 MISSING (IN REF. 2).
FT CONFLICT 27 27 P -> A (IN REF. 2).
FT CONFLICT 1352 1352 T -> S (IN REF. 2).
SQ SEQUENCE 1408 AA; 150660 MW; 569DA270A9C7840 CRC64;

Query Match 7.5%; Score 625.5; DB 1; Length 1408;
Best Local Similarity 24.2%; Pred. No. 1.8e-31;
Matches 186; Conservative 76; Mismatches 257; Indels 249; Gaps 30;

Qy 853 GANPLYDCNMQLSDWVSEYKEPGIARC--AGPGEMADKLLLTTPSKRFTCGQPV--- 907
Db 291 GCDPVHKGCD-----RPGECERPGWRGPLCNECMVTPGCKHSGCNGSAWKC 337
Qy 908 --DVN---ILAKCNPLCS-----NPCKNDGTCNSDPVDFYRCTCPYGFQDQDVPIHAC 957
Db 338 VCDTNWGIIL--CDQDLNFCGTHEPCKHGTCENTAPDKRYCTCAEGLSGEQEIVEHPC 395
Qy 958 ISNPCKHGGTCHLK-----EGE----- 974
Db 396 ATRPCRNCGTCTLKTSNRTQAQVYRTSHGRSNMGRPVRRSSMSRLDHLRPEGALNGSS 455
Qy 975 -----EDGFWICADGFECEGNEVNDDEDNCCNNSTVCDGINNYTC 1018
Db 456 SSGLVSLGSLQQLQQLAPDFTCDCAAGWGTPTCEINIDECAGGPECHGGTCDLIGGPRC 515
Qy 1019 LCPPEYTGECEKLDLC----- 1036
Db 516 ECPPEWHGDVQVDVNECEAPHAGIAANALLTTTATAIGSNLSSTALLAALTSVAST 575
Qy 1037 AQDLNCPQHDSKILTPKGFCDCTPGYGEHCIDFDQDQDNCKNGAHCDAVNGYTC 1096
Db 576 SLAIGPCINAKERNQPGSFACICEKGWGVTCENLDDCV-GQCRNGATCIDLVNDYRC 634
Qy 1097 ICPEGYSLGFCFSPMVLPRSPCDNFDQNGAQCIIVRINEPICQLPGYQGEKCEKV 1156
Db 635 ACASGPTGRDCETD-----IDECATSPCRNGGECVDMVGKFNICPLGYSGLCEEA- 686
Qy 1157 SVNFEINKESYLQIPSAKVRPQTNTLQIATDEDSGILLYKGDKDHIAVELYGRVRASYD 1216
Db 687 -----KENCTPSPCL-----EGHCLNTPGYCHCPD-----RA--- 716
Qy 1217 TGSH--PASAIYSVETINDGNFHVIELLALDQSLSLVDGGNKPIITNLKSQSTLNFDS 1274
Db 717 -GRKCEQLRPLCSQPPCNEGCFANVSL-----ATSATTTTTTTTATTRKMAKP 765

Qy 1275 LVVGGMPGKSNVASLRQAPGQNGTSHFGICRNLYINSELQDFQKVPK-OTGILPGCE--- 1330
Db 766 ---SGLP-----CSGHGSC---MSDVGTFCCKHVHTGTGTF--CEHNL 800
Qy 1331 -PCHKKVCAH3TCQPSSAGFTCECQEGWMLPCDQRTNDPCLGNKCVH-GTCL---PIN 1385
Db 801 NECSNPNCRRNGGICLDGDDFTCECMGWTGKRCSEATG-CYAGQCQNGGTCMPGAPDK 859
Qy 1386 AFSYSCKCLEHGGVLCDEEEDLFNCPQAIKCKHKGKRLSGLGPYCECSSGYTGDSCDR 1445
Db 860 ALQPHRCAPGWTGLFCAEAID---QCRGQPCNHGGTCESGAGWFRVCVCAQGGSGPDC-- 914
Qy 1446 EISRCGERIRDIYQKQGYAACQTTKVSRLECRGGCA-----GGQCC 1488
Db 915 -----RI-----NVNECSPPCQCGGATCIDGIGGYSC 941

RESULT 13

LI12_CAEEL

ID LI12_CAEEL STANDARD; PRT: 1429 AA.

AC P14585;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE LIN-12 PROTEIN PRECURSOR.

GN LIN-12 OR R107.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=88334747; PubMed=3419531;

RA Yochem J., Weston K., Greenwald I.;

RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane

protein with overall similarity to Drosophila Notch.";

RL Nature 335:547-550(1988).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Sims M., Smalldo N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,

RA Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans.";

RL Nature 368:32-33(1994).

CC -!- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT

CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12

CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES

CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.

CC -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.

CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

[illegible]

DR EMBL; L07045; AAA30045.1; -.
DR HSSP; P00740; 1IXA.
DR INTERPRO; IPR000088; -.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000859; -.
DR INTERPRO; IPR001438; -.

Query Match 6.6%; Score 549.5; DB 1; Length 570;
Best Local Similarity 39.7%; Pred. No. 3.2e-27;
Matches 96; Conservative 33; Mismatches 98; Indels 15; Gaps 5;

```
Qy 918 CLSNPCKNDGTNSDPVDFYRCTCPYFGKQCDVPIHACISNPCKHGGTCHLKEGEEDG 977
    | ||| || | : | : | : | : | : | : | : | : | : | : | : |
Db 180 CTPNPCLNGATC-VDQVNDYQICAPGFTGNCETDIDECASAPCRNGGAC---VDQVNG 235

Qy 978 FWCICADGFECEGENCEVNVDDCEDNDCENNSTCVDGINNYTCLCPPEYTGELCEEKLD FCA 1037
    : | | | | | | : : | | | | | | : | | | | | | | : | |
Db 236 YTCNCIPGFGNVNCEENNIECASIPCLNGGICVDGINQFACTCLPGYTGLCETDINECA 295

Qy 1038 QDLNPQCQHDSKCLLT PKGFKCDCTPGYVGEHCIDIFDDCQDNKCKGAHCTDAVNGYTCI 1097
    : ||| : | : ||| : | : : : : | | | | | : | : | |
Db 296 S--SPQNGGSCCTDAVNRYTCDRCRAGFTGSCNCTNIECASSPCLNGGSCLDGVDGYVCQ 353

Qy 1098 CPEGYSGLPFCFSPPMWLPRTSPCDNFDCQNGAQCVIRINEPICQLPGYQGEKCEKLVS 1157
    : | : | | : | : | | : | : | : | : | : | : | : | : |
Db 354 CLPNYTYTHCISIS-----LDACASLPQNGGVCTNVGGDVYCECLPGYTGINCE--ID 404
```

Qy 1158 VN 1159

Db 405 IN 406

RESULT 15

GLP1_CAEL

ID GLP1_CAEL STANDARD; PRT; 1295 AA.
 AC P13508;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GLP-1 PROTEIN PRECURSOR.
 GN GLP-1 OR EMB-33 OR F02A9.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=89336787; PubMed=2758466;
 RA Yochem J., Greenwald I.;
 RT "glp-1 and lin-12, genes implicated in distinct cell-cell
 RT interactions in C. elegans, encode similar transmembrane proteins.";
 RL Cell 58:553-563(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jler M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Sims M., Smaison N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP DELETION OF 1174-1295.
 RX MEDLINE=91351288; PubMed=1881436;
 RA Mango S.E., Maine E.M., Kimble J.;
 RT "Carboxy-terminal truncation activates glp-1 protein to specify
 RT vulval fates in Caenorhabditis elegans.";
 RL Nature 352:811-815(1991).
 RN [4]
 RP CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.
 RX MEDLINE=93354444; PubMed=8350921;
 RA Roehl H., Kimble J.;
 RT "Control of cell fate in C. elegans by a GLP-1 peptide consisting
 RT primarily of ankyrin repeats.";
 RL Nature 364:632-635(1993).
 RN [5]
 RP FUNCTION.
 RX MEDLINE=94208066; PubMed=8156602;
 RA Mello C.C., Draper B.W., Priess J.R.;
 RT "The maternal genes apx-1 and glp-1 and establishment of
 RT dorsal-ventral polarity in the early C. elegans embryo.";
 RL Cell 77:95-106(1994).
 CC -1- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE
 CC BLASTOMERES, ABA AND APA. PROPER SIGNALING BY GLP-1 INDUCES ABA
 CC DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA
 CC DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE
 CC ESTABLISHMENT THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: ACTS ON ABP DEVELOPMENT DURING 4-CELL AND
 CC 12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL

CC STAGES.

CC -1- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.
 CC -1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M25580; AAA28058.1; -.
 DR EMBL; Z19555; CAA79620.1; -.
 DR EMBL; Z29116; CAA79620.1; JOINED.
 DR EMBL; Z29116; CAA82373.1; -.
 DR EMBL; Z19555; CAA82373.1; JOINED.
 DR PIR; A32901; A32901.
 DR HSSP; P00740; L1XA.
 DR WORMPEP; F02A9.6; CE00237.
 DR INTERPRO; IPR000152; -.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR000800; -.
 DR INTERPRO; IPR001881; -.
 DR INTERPRO; IPR002110; -.
 DR PFAM; PF00008; EGF; 10.
 DR PFAM; PF00023; ank; 4.
 DR PFAM; PF00066; notch; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 2.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 8.
 DR PROSITE; PS01187; EGF_CA; 1.
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 1295 GLP-1 PROTEIN.
 FT DOMAIN 16 764 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 765 786 POTENTIAL.
 FT DOMAIN 787 1295 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 493 607 3 X LIN/NOTCH REPEATS.
 FT DOMAIN 988 1133 6 X ANK MOTIF REPEATS.
 FT DOMAIN 19 58 EGF-LIKE 1.
 FT DOMAIN 117 152 EGF-LIKE 2.
 FT DOMAIN 154 190 EGF-LIKE 3.
 FT DOMAIN 190 230 EGF-LIKE 4.
 FT DOMAIN 232 269 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 271 308 EGF-LIKE 6.
 FT DOMAIN 316 359 EGF-LIKE 7.
 FT DOMAIN 369 406 EGF-LIKE 8.
 FT DOMAIN 407 443 EGF-LIKE 9.
 FT DOMAIN 446 479 EGF-LIKE 10.
 FT REPEAT 493 527 LIN/NOTCH 1.
 FT REPEAT 528 568 LIN/NOTCH 2.
 FT REPEAT 569 608 LIN/NOTCH 3.
 FT REPEAT 915 946 ANK MOTIF 1.
 FT REPEAT 947 987 ANK MOTIF 2.
 FT REPEAT 988 1019 ANK MOTIF 3.
 FT REPEAT 1020 1056 ANK MOTIF 4.
 FT REPEAT 1057 1098 ANK MOTIF 5.
 FT REPEAT 1099 1133 ANK MOTIF 6.
 FT DISULFID 23 35 BY SIMILARITY.
 FT DISULFID 29 46 BY SIMILARITY.
 FT DISULFID 48 57 BY SIMILARITY.
 FT DISULFID 121 131 BY SIMILARITY.
 FT DISULFID 126 140 BY SIMILARITY.
 FT DISULFID 142 151 BY SIMILARITY.
 FT DISULFID 158 169 BY SIMILARITY.
 FT DISULFID 163 178 BY SIMILARITY.
 FT DISULFID 180 189 BY SIMILARITY.

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:19:59 ; Search time 559.88 Seconds
(without alignments)
319.251 Million cell updates/sec

Title: US-09-540-245A-2
Perfect score: 8316
Sequence: 1 MRGVGVQWMLSLSLGLVLAIL.....SSFVDEVEKVVKCGCTRCVS 1525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8316	100.0	1525	4	Q9Y5Q7	Q9y5q7 homo sapien
2	8265	99.4	1529	4	O94813	O94813 homo sapien
3	8257	99.3	1521	4	O95710	O95710 homo sapien
4	8095	97.3	1521	11	Q9R1B9	Q9r1b9 mus musculu
5	5728	68.9	1523	11	O88280	O88280 rattus norv
6	5714.5	68.7	1523	4	O75094	O75094 homo sapien
7	5703	68.6	1523	11	Q9WVB4	Q9wvb4 mus musculu
8	5597	67.3	1534	4	O75093	O75093 homo sapien
9	5578.5	67.1	1531	11	O88279	O88279 rattus norv
10	5538.5	66.6	1531	11	Q9WVB5	Q9wvb5 mus musculu
11	5530	66.5	1025	11	Q9Z166	Q9z166 mus musculu
12	5497	66.1	1530	11	Q9WUG5	Q9wug5 rattus norv
13	3934	47.3	796	11	Q9WVC1	Q9wvc1 rattus norv
14	3588	43.1	1504	5	Q9V7F9	Q9v7f9 drosophila
15	3586	43.1	1504	5	Q9XYV4	Q9xyv4 drosophila
16	3517	42.3	1480	5	Q9V7F8	Q9v7f8 drosophila
17	3071	36.9	850	4	O95804	O95804 homo sapien
18	2703.5	32.5	1440	5	Q20204	Q20204 caenorhabdi
19	2282.5	27.4	664	4	Q9U1L7	Q9u1l7 homo sapien

20	1318	15.8	333	4	Q9UFH5	Q9ufh5 homo sapien
21	1115.5	13.4	530	5	Q24526	Q24526 drosophila
22	818.5	9.8	2704	5	O97458	O97458 drosophila
23	818	9.8	2634	5	Q9W4T8	Q9w4t8 drosophila
24	790	9.5	2653	5	Q25253	Q25253 lucilia cup
25	768	9.2	1218	11	Q9QXX0	Q9qxx0 mus musculu
26	768	9.2	1219	11	Q63722	Q63722 rattus norv
27	766	9.2	1218	4	Q15816	Q15816 homo sapien
28	766	9.2	1218	4	O14902	O14902 homo sapien
29	766	9.2	1227	4	P78504	P78504 homo sapien
30	761	9.2	1218	4	O15122	O15122 homo sapien
31	757	9.1	2352	5	O61240	O61240 halocynthia
32	748.5	9.0	2447	13	O13149	O13149 fugu rubrip
33	736.5	8.9	1193	13	Q90819	Q90819 gallus gall
34	736	8.9	1203	11	Q06008	Q06008 mus musculu
35	735	8.8	2470	11	O35516	O35516 mus musculu
36	732	8.8	2471	11	Q9QW30	Q9qw30 rattus sp.
37	731	8.8	2531	5	O16004	O16004 lytechinus
38	719.5	8.7	2281	4	Q9UPL3	Q9upl3 homo sapien
39	719.5	8.7	2321	4	Q9Y6L8	Q9y6l8 homo sapien
40	719.5	8.7	2321	4	Q9UM47	Q9um47 homo sapien
41	713.5	8.6	1212	13	O42347	O42347 gallus gall
42	711.5	8.6	2319	11	Q9R172	Q9r172 rattus norv
43	697	8.4	1254	13	Q9YHU2	Q9yhu2 brachydanio
44	691.5	8.3	1964	11	O35442	O35442 mus musculu
45	682.5	8.2	1238	4	Q9Y219	Q9y219 homo sapien

ALIGNMENTS

RESULT 1
Q9Y5Q7
ID Q9Y5Q7 PRELIMINARY; PRT: 1525 AA.
AC Q9Y5Q7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT2.
GN SLIT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99200389; PubMed=10102266;
RA Wang K.H., Brose K., Arnott D., Kidd T., Goodman C.S., Henzel W.,
RA Tessier-Lavigne M.;
RT "Biochemical purification of a mammalian slit protein as a positive
RT regulator of sensory axon elongation and branching.";
RL Cell 96:771-784(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
DR EMBL; AF133270; AAD25539.1; -.
DR HSSP; P00743; 1CCF.
DR INTERPRO; IPR0003152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.

DR INTERPRO: IPR002272; -.
DR PFAM: PF00008; EGF; 9.
DR PFAM: PF00054; laminin_G; 1.
DR PFAM: PF00560; LRR; 18.
DR PFAM: PF01462; LRRNT; 4.
DR PFAM: PF01463; LRRCT; 4.
DR PRINTS: PR01143; FSHRECEPTOR.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 7.
DR PROSITE: PS01187; EGF_CA; 2.
DR PROSITE: PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1525 AA; 169394 MW; 8A81CDE34EF06A73 CRC64;

Query Match 100.0%; Score 8316; DB 4; Length 1525;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRGVGQMLSLSLGLVLAILNKVAPQACQSCSGSTVDCHGLALRSVPRNIPRTERL 60
1 MRGVGQMLSLSLGLVLAILNKVAPQACQSCSGSTVDCHGLALRSVPRNIPRTERL 60
61 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRNRLNLQLPPE 120
61 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRNRLNLQLPPE 120
121 LLFLGTAKLYRLDLSENQIAIPKAFRGAVDIKNLQLDYNIQSCIEDGAFRALRDLEVL 180
121 LLFLGTAKLYRLDLSENQIAIPKAFRGAVDIKNLQLDYNIQSCIEDGAFRALRDLEVL 180
181 TLNNNNITRLSVASFNHPKLRTRFLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPS 240
181 TLNNNNITRLSVASFNHPKLRTRFLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPS 240
241 HLRGHNAEVQKREFVCSDEEEHGSFMAPSCSVLHCPAACTCSNNIVDCRGKLTETPT 300
241 HLRGHNAEVQKREFVCSDEEEHGSFMAPSCSVLHCPAACTCSNNIVDCRGKLTETPT 300
301 NLPETITEIRLEQNTIKVIPPQAFSPYKLRRLIDLSNNQISELAPDAFQGLSLNSLVLY 360
301 NLPETITEIRLEQNTIKVIPPQAFSPYKLRRLIDLSNNQISELAPDAFQGLSLNSLVLY 360
361 GNKITELPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLDNKLQTIAGTF 420
361 GNKITELPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLDNKLQTIAGTF 420
421 SPLRAIQTMLHAQNPFICDCHLKLWADLYHTNPIETSGARCTSPRRLANKRIGQIKSKF 480
421 SPLRAIQTMLHAQNPFICDCHLKLWADLYHTNPIETSGARCTSPRRLANKRIGQIKSKF 480
481 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKPIEPIQYTAELRLNNE 540
481 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKPIEPIQYTAELRLNNE 540
541 FTVLEATGIFKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLENVQHKMFG 600
541 FTVLEATGIFKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLENVQHKMFG 600
601 LESLKTMLRSNRITCVGNDSFISLSSVRLSLYDNIQTTVAPGAFDILHSLSTLNLN 660
601 LESLKTMLRSNRITCVGNDSFISLSSVRLSLYDNIQTTVAPGAFDILHSLSTLNLN 660
661 PFCNCYLAWLGEWLKRRIVTGNPRCQKPYFLKEIPIDQVAIQDFTCDGNDNCSPL 720
661 PFCNCYLAWLGEWLKRRIVTGNPRCQKPYFLKEIPIDQVAIQDFTCDGNDNCSPL 720
721 SRCPECTCLDTVVRCSNKGKLVLPKGPIDRVTELYDNGQFTLVPKELSNYKHLTLIDL 780
721 SRCPECTCLDTVVRCSNKGKLVLPKGPIDRVTELYDNGQFTLVPKELSNYKHLTLIDL 780

781 SNNRISTLSNQSFNSMTQLTLILSYNRLRCIPPTFDGLKSLRLSLHGNDISVVPEGA 840
781 SNNRISTLSNQSFNSMTQLTLILSYNRLRCIPPTFDGLKSLRLSLHGNDISVVPEGA 840
841 FNDLSALSHLAIGANPLYCDCNMQWLSWVSEYKEPGIARCAPGEMADKLLTTPSKK 900
841 FNDLSALSHLAIGANPLYCDCNMQWLSWVSEYKEPGIARCAPGEMADKLLTTPSKK 900
901 FTCQGPVDVNIKACNPCLSNPCKNDGTCNSDPVDFYRCTCPYFGKQDCDVIHACISN 960
901 FTCQGPVDVNIKACNPCLSNPCKNDGTCNSDPVDFYRCTCPYFGKQDCDVIHACISN 960
961 PCKHGGTCHLKEGEEDGFWCICADGFECEVNVDDCEDNDCENNSTCVDGINNYTCLC 1020
961 PCKHGGTCHLKEGEEDGFWCICADGFECEVNVDDCEDNDCENNSTCVDGINNYTCLC 1020
1021 PPEYTGELCEKLDPCAQDLNCPQHDSKILTPKGFCDCTPGYVGEHCDIDFDDCQDNK 1080
1021 PPEYTGELCEKLDPCAQDLNCPQHDSKILTPKGFCDCTPGYVGEHCDIDFDDCQDNK 1080
1081 CKNGAHCDAVNGYTCICPEGISGLCFECSPPMVLPRTPSPCNDFCQNGAQCIVRINEPI 1140
1081 CKNGAHCDAVNGYTCICPEGISGLCFECSPPMVLPRTPSPCNDFCQNGAQCIVRINEPI 1140
1141 CQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGD 1200
1141 CQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGD 1200
1201 HIAVELYRGRVRSYDTGSHPASAIYSVETINDGNFHVHALLDQSLSLSDGNGPKII 1260
1201 HIAVELYRGRVRSYDTGSHPASAIYSVETINDGNFHVHALLDQSLSLSDGNGPKII 1260
1261 TNLKQSTLNFDSPLYVGGMPKSNVASLRQAPGQNGTSFHGCIIRNLYINSELQDFQKVP 1320
1261 TNLKQSTLNFDSPLYVGGMPKSNVASLRQAPGQNGTSFHGCIIRNLYINSELQDFQKVP 1320
1321 MQTGILPGCEPCHKKVCAHGTCPSSQAGFTCECEGWMGPLCQRTNDPCLGNKCVHGT 1380
1321 MQTGILPGCEPCHKKVCAHGTCPSSQAGFTCECEGWMGPLCQRTNDPCLGNKCVHGT 1380
1381 CLPINAFTSYSCKLEGGHGVLCDEEEDLFNCPQAIKCKHKGKRLSGLGQPYCECSSGYTG 1440
1381 CLPINAFTSYSCKLEGGHGVLCDEEEDLFNCPQAIKCKHKGKRLSGLGQPYCECSSGYTG 1440
1441 DSCDREISCRGERIRDIYQKQGYAACQTTKVSRLCRGGCAGGQCCGLRSKRRKYSF 1500
1441 DSCDREISCRGERIRDIYQKQGYAACQTTKVSRLCRGGCAGGQCCGLRSKRRKYSF 1500
1501 ECTDGSSFVDEKVVKCGCTRCVS 1525
1501 ECTDGSSFVDEKVVKCGCTRCVS 1525

RESULT 2

O94813

ID O94813 PRELIMINARY; PRT; 1529 AA.

AC O94813;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE SLIT-2 PROTEIN.

GN SLIT-2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99033071; PubMed=9813312;

RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;

RT "Cloning and expressions of three mammalian homologues of Drosophila
slit suggest possible roles for Slit in the formation and maintenance
of the nervous system.";

RL Brain Res. Mol. Brain Res. 62:175-186(1998).
DR EMBL; AB017168; BAA35185.1; -.
DR HSSP; P00743; 1CCF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002272; -.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_6; 1.
DR PFAM; PF00560; LRR; 18.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
PRINTS; PRO1143; FSHRECEPTOR.
PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1529 AA; 169868 MW; 5D19CC5E7FD461BA CRC64;

Query Match 99.4%; Score 8265; DB 4; Length 1529;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1520; Conservative 1; Mismatches 0; Indels 12; Gaps 2;

Qy 1 MRGVGQMLSLSLGLVLAILNKVAPQACPAQCSGSGTVDCHGLALRSVPRNIPRNTLRL 60
Db 1 MRGVGQMLSLSLGLVLAILNKVAPQACPAQCSGSGTVDCHGLALRSVPRNIPRNTLRL 60
Qy 61 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRLNRNLQLFPE 120
Db 61 DLNGNNITRITKTDFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRLNRNLQLFPE 120
Qy 121 LLPLGTAKLYRLDLSNQIQAIPKAFRGAVDIKNLQLDYNIQSCIEDGAFRALRDLEVL 180
Db 121 LLPLGTAKLYRLDLSNQIQAIPKAFRGAVDIKNLQLDYNIQSCIEDGAFRALRDLEVL 180
Qy 181 TLNNNNITRLSVASFNMHMKLRTFRLHSNNLYCDCHLAWLSDWLKRPVGLYTQCMGPS 240
Db 181 TLNNNNITRLSVASFNMHMKLRTFRLHSNNLYCDCHLAWLSDWLKRPVGLYTQCMGPS 240
Qy 241 HLRGHNVAEVQKREFVCSDEEGHQSFMAPSCSVLHCAACTCSNNIVDCRGKGLTEIPT 300
Db 241 HLRGHNVAEVQKREFVCS---GHQSFMAPSCSVLHCAACTCSNNIVDCRGKGLTEIPT 296
Qy 301 NLPETITEIRLEQNTIKVIPPAGAFSPYKLRRLIDLSNNQISELAPDAFQGLRSLNSLVLY 360
Db 297 NLPETITEIRLEQNTIKVIPPAGAFSPYKLRRLIDLSNNQISELAPDAFQGLRSLNSLVLY 356
Qy 361 GNKITELPKSLFEGLFSLQLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLTQIAKGT 420
Db 357 GNKITELPKSLFEGLFSLQLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLTQIAKGT 416
Qy 421 SPLRAIQTMHLAQNPFICDCHLKLWADLYHTNPIETSGARCTSPRRANKRIGQIKSKF 480
Db 417 SPLRAIQTMHLAQNPFICDCHLKLWADLYHTNPIETSGARCTSPRRANKRIGQIKSKF 476
Qy 481 RCS-----GTEDYRSKLSGDCFADLACPEKRCCEGTVDVCSNQLNKIPEHIPQYTA 532
Db 477 RCSAEQYFIPGTEDYRSKLSGDCFADLACPEKRCCEGTVDVCSNQLNKIPEHIPQYTA 536
Qy 533 ELRLNNEFTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLN 592
Db 537 ELRLNNEFTVLEATGIFKKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLN 596

Qy 593 VQHKMFKGLSGLKTLMLRSNRITCVGNDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSL 652
Db 597 VQHKMFKGLSGLKTLMLRSNRITCVGNDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSL 656
Qy 653 STLNLNLPNFCNCYLAWLGEWLKRRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGN 712
Db 657 STLNLNLPNFCNCYLAWLGEWLKRRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGN 716
Qy 713 DDNSCSPLSRCPTCTCLDTVVRCNSKGLKVLPGKIPRDVTELYLDGNQFTLVPKELSNY 772
Db 717 DDNSCSPLSRCPTCTCLDTVVRCNSKGLKVLPGKIPRDVTELYLDGNQFTLVPKELSNY 776
Qy 773 KHLTLIDLSNNRISTLSNQSFNMTQLTLILSYNRLRCIPPTFDGLKSLRLSLHGDND 832
Db 777 KHLTLIDLSNNRISTLSNQSFNMTQLTLILSYNRLRCIPPTFDGLKSLRLSLHGDND 836
Qy 833 ISVVPBGAFNQLSALSHLAIGANPLYCDGNQWLDWVSEYKEPGIARCAAGPGEADKL 892
Db 837 ISVVPBGAFNQLSALSHLAIGANPLYCDGNQWLDWVSEYKEPGIARCAAGPGEADKL 896
Qy 893 LLTPSKKFTCQGPVDNVLAKCNPLSNPKNDGTCNSDPVDFYRCTCPYFGKQDCDV 952
Db 897 LLTPSKKFTCQGPVDNVLAKCNPLSNPKNDGTCNSDPVDFYRCTCPYFGKQDCDV 956
Qy 953 PIHACISNPKKHGGTCHLKEGEEDGFWICADGFEGENCEVNVDDCEDNDCENNSTCVDG 1012
Db 957 PIHACISNPKKHGGTCHLKEGEEDGFWICADGFEGENCEVNVDDCEDNDCENNSTCVDG 1016
Qy 1013 INNYTCLCPPEYTGELCEEKLDFAQDLNCPQHDSKICILTPKGFKCDCTPGYVGEHCDID 1072
Db 1017 INNYTCLCPPEYTGELCEEKLDFAQDLNCPQHDSKICILTPKGFKCDCTPGYVGEHCDID 1076
Qy 1073 FDDCQDNKCKNGAHTDAVNGYTICPEGYSGLFCEFSPPMVLPRTPSPDNFDCQNGAQC 1132
Db 1077 FDDCQDNKCKNGAHTDAVNGYTICPEGYSGLFCEFSPPMVLPRTPSPDNFDCQNGAQC 1136
Qy 1133 IVRINEPICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNTLQIATDEDSGI 1192
Db 1137 IVRINEPICQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNTLQIATDEDSGI 1196
Qy 1193 LLYKGDKDHIAVELYRGRVRSYDTGSHPSAISVETINDGNFHVIELLALDQSLSLSV 1252
Db 1197 LLYKGDKDHIAVELYRGRVRSYDTGSHPSAISVETINDGNFHVIELLALDQSLSLSV 1256
Qy 1253 DGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRAPQNGTSTFHGICIRNLYNSE 1312
Db 1257 DGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVASLRAPQNGTSTFHGICIRNLYNSE 1316
Qy 1313 LQDFQKVPMTGILPGCEPCHKVKVAHGTCPQSSQAGFTCECQEGWMLPCDQRTNDPCL 1372
Db 1317 LQDFQKVPMTGILPGCEPCHKVKVAHGTCPQSSQAGFTCECQEGWMLPCDQRTNDPCL 1376
Qy 1373 GNKCVHGTCLPINAFSYCKCLEGGHGLCDEEDLFPNCPQAIKCKHGCRLSGLGQPYC 1432
Db 1377 GNKCVHGTCLPINAFSYCKCLEGGHGLCDEEDLFPNCPQAIKCKHGCRLSGLGQPYC 1436
Qy 1433 ECSSGYTGDSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLR 1492
Db 1437 ECSSGYTGDSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLR 1496
Qy 1493 SKRRKYSPECTDGSSFVDEVEKVVKCGCTRCVS 1525
Db 1497 SKRRKYSPECTDGSSFVDEVEKVVKCGCTRCVS 1529

RESULT 3
O95710
ID O95710 PRELIMINARY; PRT; 1521 AA.
AC O95710;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE NEUROGENIC EXTRACELLULAR SLIT PROTEIN SLIT2.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY, BRAIN;
RX MEDLINE=99279238; PubMed=10349621;
RA Holmes G.P., Negus K., Burridge L., Raman S., Algar E., Yamada T.,
RA Little M.H.;
RT "Distinct but overlapping expression patterns of two vertebrate slit
RT homologs implies functional roles in CNS development and
RT organogenesis.";
RL Mech. Dev. 79:57-72(1998).
DR EMBL; AF055585; AAD04309.1; -.
DR HSSP; P00743; 1CCF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002272; -.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 18.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PR01143; FSHRECEPTOR.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1521 AA; 168947 MW; C05A0DF7D78C48C9 CRC64;

Query Match 99.3%; Score 8257; DB 4; Length 1521;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1515; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

Qy 1 MRGVGQWMLSLSLGLVLAAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL 60
Db 1 MRGVGQWMLSLSLGLVLAAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL 60
61 DLNGNNITRITKTDFAGLRHLRLVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLPPE 120
Db 61 DLNGNNITRITKTDFAGLRHLRLVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLPPE 120
121 LLFLGTAKLYRLDSENQIAIPKAFRGAVDIKNLQDYNQISCIEDGAFRALRDLEVL 180
Db 121 LLFLGTAKLYRLDSENQIAIPKAFRGAVDIKNLQDYNQISCIEDGAFRALRDLEVL 180
181 TLNNNNITRLSVASFNMPLKRLTFRLHSNNLYCCHLAWLSDWLKRRPRVGLYTQCMGPS 240
Db 181 TLNNNNITRLSVASFNMPLKRLTFRLHSNNLYCCHLAWLSDWLKRRPRVGLYTQCMGPS 240
241 HLRGHNAEVQKREFVCSDEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 300
Db 241 HLRGHNAEVQKREFVCS---GHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 296
301 NLPETITEIRLEQNTIKVIPPAGAFSPYKLRRLIDLSNNQISELPADAFGLRSLNSLVLY 360
Db 297 NLPETITEIRLEQNTIKVIPPAGAFSPYKLRRLIDLSNNQISELPADAFGLRSLNSLVLY 356
361 GNKITEPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLTQIAKGT 420
Db 357 GNKITEPKSLFEGFLSLQLLLNANKINCLRVDAFQDLHNLNLSLYDNKLTQIAKGT 416

Qy 421 SPLRAIQTMHLAQNPFFICDCHLKLWADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKF 480
Db 417 SPLRAIQTMHLAQNPFFICDCHLKLWADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKF 476
481 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNE 540
Db 477 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPEHIPQYTAELRLNNE 536
541 FTVLEATGIFKKLPQLRKINFSNNKITDIEEAGFEGASGVNEILLTSNRLNVQHKMFKG 600
Db 537 FTVLEATGIFKKLPQLRKINFSNNKITDIEEAGFEGASGVNEILLTSNRLNVQHKMFKG 596
601 LESLKTMLRSNRITCVGNDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLNLAN 660
Db 597 LEKPNMLRSNRITCVGNDSFIGLSSVRLSLYDNQITTVAPGAFDTLHSLSTLNLNLAN 656
661 PFNCNCYLAWLGEWLRRKRIVTGNPRCKPYFLKEIPIQDVAIQDFTCDDGNDNDSCSPL 720
Db 657 PFNCNCYLAWLGEWLRRKRIVTGNPRCKPYFLKEIPIQDVAIQDFTCDDGNDNDSCSPL 716
721 SRCPTCTCLDTVVRCNSNKGKLVLPKGIIPDVTELYLDGNQFTLVPKELSNYKHLTLIDL 780
Db 717 SRCPTCTCLDTVVRCNSNKGKLVLPKGIIPDVTELYLDGNQFTLVPKELSNYKHLTLIDL 776
781 SNNRISTLSNQSFNNMTQLLLTLILSYNRLRCIPRPTFDGLKSLRLLSLHGNDISVVPGEA 840
Db 777 SNNRISTLSNQSFNNMTQLLLTLILSYNRLRCIPRPTFDGLKSLRLLSLHGNDISVVPGEA 836
841 FNDLSALSHLAIGANPLYCDNCNQWLDWVSEYKEPGIARCAAGPGEMADKLLLTTPSKK 900
Db 837 FNDLSALSHLAIGANPLYCDNCNQWLDWVSEYKEPGIARCAAGPGEMADKLLLTTPSKK 896
901 FTCQGPVDVNI LAKCNPCLSNPKNDGTCSNDPVDYRCTCPYGFKGQDCDVPIHACISN 960
Db 897 FTCQGPVDVNI LAKCNPCLSNPKNDGTCSNDPVDYRCTCPYGFKGQDCDVPIHACISN 956
961 PCKHGGTCHLKEGEEDGFWCICADGFECECENNVDDCEDNDCENNSTCVDGINNYTCLC 1020
Db 957 PCKHGGTCHLKEGEEDGFWCICADGFECECENNVDDCEDNDCENNSTCVDGINNYTCLC 1016
1021 PPEYTGELCEBKLDPCAQLNCPQHDSKCLITPKGFKCDCTPGYVGEHCDIDFDDCQDNK 1080
Db 1017 PPEYTGELCEBKLDPCAQLNCPQHDSKCLITPKGFKCDCTPGYVGEHCDIDFDDCQDNK 1076
1081 CKNGAHCITDAVNGYTCICPEYSGLFCEFSPPMVLPRTPCNDNFCQNGAQICVRINEPI 1140
Db 1077 CKNGAHCITDAVNGYTCICPEYSGLFCEFSPPMVLPRTPCNDNFCQNGAQICVRINEPI 1136
1141 CQCLPGYQGEKCEKLSVNFINKESYLQIPSAKVRPQNTILQIATDEDSGILLYKGDKD 1200
Db 1137 CQCLPGYQGEKCEKLSVNFINKESYLQIPSAKVRPQNTILQIATDEDSGILLYKGDKD 1196
1201 HIAVELYRGVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSDVGGNPKII 1260
Db 1197 HIAVELYRGVRASYDTGSHPASAIYSVETINDGNFHIVELLALDQSLSLSDVGGNPKII 1256
1261 TNLKSQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHHGICRNLINSELQDFQKVP 1320
Db 1257 TNLKSQSTLNFDSPLYVGGMPGKSNVASLRQAPGQNGTSFHHGICRNLINSELQDFQKVP 1316
1321 MQTGILPGCEPCHKKVCAHGTCPQSSQAGFTCEQEGWMPGLCDQRTNDPCLGNKCVHGT 1380
Db 1317 MQTGILPGCEPCHKKVCAHGTCPQSSQAGFTCEQEGWMPGLCDQRTNDPCLGNKCVHGT 1376
1381 CLPINFASYSCKLEGGHGVLCDEEEDLFNPQCAIKCKHKGKRLSGLGQPYCESSGYTG 1440
Db 1377 CLPINFASYSCKLEGGHGVLCDEEEDLFNPQCAIKCKHKGKRLSGLGQPYCESSGYTG 1436
1441 DSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLSRKRKYSF 1500
Db 1437 DSCDREISCRGERIRDYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLSRKRKYSF 1496

Qy 1501 ECTDGSSFVDEVEKVVKCGCTRCVS 1525

Db 1497 ECTDGSSFVDEVEKVVKCGCTRCVS 1521

RESULT 4

Q9R1B9

ID Q9R1B9 PRELIMINARY; PRT: 1521 AA.

AC Q9R1B9;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE SLIT2.

GN SLIT2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

[1]

SEQUENCE FROM N.A.

RC STRAIN=SWISS WEBSTER/ICR;

RA Yuan W., Zhou L., Chen J.-h., Wu J.Y., Rao Y., Ornitz D.M.;

RT "The mouse SLIT family: Secreted ligands for Robo expressed in
RT patterns that suggest a role in morphogenesis and axon guidance.";

RL Dev. Biol. 0:0-0(1999).

DR EMBL; AF144628; AAD44759.1; -.

DR HSSP; P00743; ICCF.

DR MGD; MGI:1315205; Slit2.

DR INTERPRO; IPR000152; -.

DR INTERPRO; IPR000359; -.

DR INTERPRO; IPR000372; -.

DR INTERPRO; IPR000483; -.

DR INTERPRO; IPR000561; -.

DR INTERPRO; IPR000742; -.

DR INTERPRO; IPR001611; -.

DR INTERPRO; IPR001791; -.

DR INTERPRO; IPR001881; -.

DR INTERPRO; IPR002272; -.

DR PFAM; PF00008; EGF; 9.

DR PFAM; PF00054; laminin_G; 1.

DR PFAM; PF00560; LRR; 18.

DR PFAM; PF01462; LRRNT; 4.

DR PFAM; PF01463; LRRCT; 4.

DR PRINTS; PR00019; LEURICHRPT.

DR PRINTS; PR01143; FSHRECEPTOR.

DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.

DR PROSITE; PS00022; EGF_1; UNKNOWN_9.

DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.

DR PROSITE; PS01186; EGF_2; 7.

DR PROSITE; PS01187; EGF_CA; 2.

SQ SEQUENCE 1521 AA; 168769 MW; 97DCA361578978E4 CRC64;

Query Match 97.3%; Score 8095; DB 11; Length 1521;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 1471; Conservative 34; Mismatches 16; Indels 4; Gaps 1;

Qy 1 MRGVGWQLSLSLGLVLAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL 60

Db 1 MSGIGWQLSLSLGLVLAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPRNIPRNTERL 60

Qy 61 DLNGNNITRITKTFAGLRHLRLVQLMENKISTIERGAFQDLKELERLRNRNLQLFPE 120

Db 61 DLNGNNITRITKTFAGLRHLRLVQLMENRISTIERGAFQDLKELERLRNRNLQLFPE 120

Qy 121 LFLGTAKLYRLDLSNQIAIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL 180

Db 121 LFLGTAKLYRLDLSNQIAIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVL 180

Qy 181 TLNNNNITRLSVASFNMHPKLRTRFLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPS 240

Db 181 TLNNNNITRLSVASFNMHPKLRTRFLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPS 240

Qy 241 HLRGHNVAEVQKREFVCSDEEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 300

Db 241 HLRGHNVAEVQKREFVCS---GHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPT 296

Qy 301 NLPETITEIRLEQNTIKVIPPAGFSPYKLRRLDLSNNQISELAPDAFQGLRSLNSLVLY 360

Db 297 NLPETITEIRLEQNSIRVIPPAGFSPYKLRRLDLSNNQISELAPDAFQGLRSLNSLVLY 356

Qy 361 GNKITELPKSLFEGFLSLQLLLNNANKINCLRVDAFQDLHNLNLSLYDNKLQTIAGKTF 420

Db 357 GNKITELPKSLFEGFLSLQLLLNNANKINCLRVDAFQDLHNLNLSLYDNKLQTIAGKTF 416

Qy 421 SPLRAIQTMLHAQNPFICDCHLKLADYLHTNPIETSGARCTSPRLANKRIGQIKSKKF 480

Db 417 SALRAIQTMLHAQNPFICDCHLKLADYLHTNPIETSGARCTSPRLANKRIGQIKSKKF 476

Qy 481 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPHIPOYTAE LRNNNE 540

Db 477 RCSGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKIPHIPOYTAE LRNNNE 536

Qy 541 FTVLEATGIFKKLPQLRKINFSNNKIIDIEGAFEGASGVNEILLTSNRLENVQHKMFKG 600

Db 537 FTVLEATGIFKKLPQLRKINFSNNKIIDIEGAFEGASGVNEILLTSNRLENVQHKMFKG 596

Qy 601 LESLKTMLRSNRITCVGNSDFIGLSSVRLSLYDNQITTVAPGAFDTHLSLSTNLNLAN 660

Db 597 LESLKTMLRSNRITCVGNSDFIGLSSVRLSLYDNQITTVAPGAFDTHLSLSTNLNLAN 656

Qy 661 PFNCNCYLAWLGEWLRRKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDNNSCSP 720

Db 657 PFNCNCYLAWLGEWLRRKRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDNNSCSP 716

Qy 721 SRCPTCTCLDTVVRCNSKGLKVLPGKIPDVTELYLDGNQFTLVPKELSNYKHLTLIDL 780

Db 717 SRCPSECTCLDTVVRCNSKGLKVLPGKIPDVTELYLDGNQFTLVPKELSNYKHLTLIDL 776

Qy 781 SNNRISTLSNQSFNSMTQLLTILSYNRLRCPPTFDGLKSLRLLSHGNDISVVPEGA 840

Db 777 SNNRISTLSNQSFNSMTQLLTILSYNRLRCPPTFDGLKSLRLLSHGNDISVVPEGA 836

Qy 841 FNDLSALSHLAIGANPLYCDNCNMQWLSDWVWSEYKEPGIARCAPGEMADKLLLTTPSKK 900

Db 837 FNDLSALSHLAIGANPLYCDNCNMQWLSDWVWSEYKEPGIARCAPGEMADKLLLTTPSKK 896

Qy 901 FTCQGPVDVNI LAKCNPCLSNPKNDGTCSNDPVDVRYCTCPYGFQDQDVPVHACISN 960

Db 897 FTCQGPVDIT IAKCNPCLSNPKNDGTCSNDPVDVRYCTCPYGFQDQDVPVHACISN 956

Qy 961 PCKHGGTCHLKEGEEGFWICADGFECEVNVDDCEDNDCENNSTCVDGNNYTCCLC 1020

Db 957 PCKHGGTCHLKEGENAGFWCTADGFECEVNVDDCEDNDCENNSTCVDGNNYTCCLC 1016

Qy 1021 PPEYTGELCEEKLDFAQDLNCPQHDSKILTPKGFCDCTPGYVGEHCDIDFDDCQDNK 1080

Db 1017 PPEYTGELCEEKLDFAQDLNCPQHDSKILTPKGFCDCTPGYVGEHCDIDFDDCQDNK 1076

Qy 1081 CKNGAHTDAVNGYTCICPEGSGLFCEFSPPMVLPTSPCDNFDCQNGAQCIVRINEPI 1140

Db 1077 CKNGAHTDAVNGYTCVCPBGYSGLFCEFSPPMVLPTSPCDNFDCQNGAQCIVRINEPI 1136

Qy 1141 CQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNTILQIATDEDSGILLYKGD 1200

Db 1137 CQCLPGYQGEKCEKLVSVNFINKESYLQIPSAKVRPQTNTILQIATDEDSGILLYKGD 1196

Qy 1201 HIAVELYRGVRASYDTGSHPSAISVETINDGNFHVIELLALDQSLSLSDVGGNPKII 1260

Db 1197 HIAVELYRGVRASYDTGSHPSAISVETINDGNFHVIELLALDQSLSLSDVGGNPKII 1256

Qy 1261 TNLKQSTLNFDSPLYVGGMPGKSNVASLRQAPQNGTSFHGICIRNLYINSELQDFQKVP 1320

Db 1257 TNLKQSTLNFDSPLYVGGMPGKSNVASLRQAPQNGTSFHGICIRNLYINSELQDFQKVP 1316

Qy 1321 MQTGILPGCECHKKVAHGTCQPSQAGFTCECQEGWMPGLCDQRTNDPCLGNKCVHGT 1380

Db 1317 MGTGILPGCEPCHKKVCAHGMCPSSQSGFTCECEGWMGLCDQRTNDPCLGKNCVHGT 1376
Qy 1381 CLPINAISYSCKLEGGHGVLCDEEEDLFNPCQAIKCKHGKRLSGVGQPYCECSSGYTG 1440
Db 1377 CLPINAISYSCKLEGGHGVLCDEEEDLFNPCQAIKCKHGKRLSGVGQPYCECNSGFTG 1436
Qy 1441 DSCDREISCRGERIRDYYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLSRKRRKYSF 1500
Db 1437 DSCDREISCRGERIRDYYQKQGYAACQTTKKVSRLECRGGCAGGQCCGPLSRKRRKYSF 1496
Qy 1501 ECTDGSFVDEVEKVVKCGCTRCVS 1525
Db 1497 ECTDGSFVDEVEKVVKCGCARCAS 1521

RESULT 5
088280
ID 088280 PRELIMINARY; PRT; 1523 AA.
AC 088280;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MEGF5.
GN MEGF5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RT like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
DR EMBL; AB011531; BAA32461.1; -.
DR HSSP; P01132; 1EGF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002400; -.
DR PFAM; PF00007; Cys_knot; 1.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 19.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PR00438; GFCYSKNOT.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1523 AA; 167766 MW; 6CE1B7AF9244478E CRC64;

Query Match 68.9%; Score 5728; DB 11; Length 1523;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1015; Conservative 223; Mismatches 265; Indels 14; Gaps 9;

Qy 11 LSLGLVLA-ILNKVAPQACPAQCSGSDVCHGLALRSVPRNIPNTERLDLGNINTR 69
Db 16 LALALALASILSGPPAACPTKTCSAASVDCHGLRAVPRGIPNAERLDLDRNNITR 75

Qy 70 ITKTDFAGLRLRLVQLMENKISTIERGAFQDLKELERLRNRLNHLQFLPELLFSTAKL 129
Db 76 ITKMDFTGLKRLRLVHLEDNQSVIERGAFQDLKELERLRNRLNHLQFLPELLFSTPKL 135
Qy 130 YRLDLSENQIAIPKAFRGAVDIKNLQLDYNQISCIEDGAFRALRDLEVLTLNNNNITR 189
Db 136 TRLDLSENQIQGIPKAFRGVTVGNLQLDNNHISCIEDGAFRALRDLEVLTLNNNNISR 195
Qy 190 LSVASFNHMPKLRTRFLSHNNLYCDCHLAWLSDWLKRPRVGLYTQCMGPSHLRGNHVAE 249
Db 196 ILVTSFNHMPKIRTLRLSHNNLYCDCHLAWLSDWLKRRTTIGQTLCAVPVHLRGSFVAD 255
Qy 250 VQKREFVCSDEEGHQSFMAPSCSV--LHCPAACTCSNNIVDCRGKLTETPNLPETIT 307
Db 256 VQKKEYVC---PGPHS-EAPACNANSLSCPSACSCSNNIVDCRGKLTETPANPEGIV 310
Qy 308 EIRLEQNTIKVIPPAGFSPIYKLRRLIDSNQISELAPDAFQGLRSLNSLVLYGNKITEL 367
Db 311 EIRLEQNSIKSIPAGAFIQYKLRRLIDSNQISELAPDAFQGLRSLNSLVLYGNKITEL 370
Qy 368 PKSLFEGFLSLQLLLNANKINCLRVDAFDLHNLNLSLYDNKLQTIAGKTFSPRLAIQ 427
Db 371 PKGLFDGLSLQLLLNANKINCLRVNTFDLHNLNLSLYDNKLQTIAGKTFSPRLAIQ 430
Qy 428 TMHLAQNPFICDCHLKLADYLTNPIETSGARCTSPRRANKRIGQIKSKKFRCSGTED 487
Db 431 TLHLAQNPFVCDCHLKLADYLDNPIETSGARCTSPRRANKRISQIKSKKFRCSGTED 490
Qy 488 YRSKLSGDCFADLACPEKCRCEGTVDSCNQKLNKIPEHIPQYTAELRNNNEFTVLEAT 547
Db 491 YRNFSSSECFMDLVCPEKCRCEGTVDSCNQKLSRIPSHLPEYTTDLRLNDNDIAVLEAT 550
Qy 548 GIFKKLPQLRKINFSNNKIITDEGAFEGASGVNEILLTNRLENVQHMKFGLSGLKTL 607
Db 551 GIFKKLPNLKINLNNRIKEVRGAFDGAAGVBLMTGNQLETHMGRMFRGLSGKLT 610
Qy 608 MLRSNRITCVGNDSFGLSVRLSLYDNQITTVAPGAFDTLHSLTNLNLNANPNCNCY 667
Db 611 MLRSNLISCVNDTFAGLSVRLSLYDNRTTISPAGFTLVSLTNLNLNANPNCNCH 670
Qy 668 LAWLGEWLAKKRIVTGNPCQPKYFLKEIPQDVAIQDFTCDGNDNDSNCSPLSRCTEC 727
Db 671 MAWLGRWLAKKRIVSGNPRCQPKYFLKEIPQDVAIQDFTC-EGNEENSQSLSPRCPEQ 729
Qy 728 TCLDVTVRCSNKGKLVLPKGPIDVTETLYLDGNTLVPKELSNYKHLTLIDLSNNRIST 787
Db 730 TCVETVVRCSNRLHTLPKGMKDVTELYLEGHNLTAVPKELSTFQRLTLIDLSNNSISM 789
Qy 788 LSNQSFNMTQLTLILSYNRLRCIPPTFDGLKSLRLSLHGNDISVVEGAFNDLSAL 847
Db 790 LTNHTFSNMSHLSTLILSYNRLRCIPVHAFNGLRSLRVTLHGNDISVVEGAFNDLSAL 849
Qy 848 SHLAIGANPLYCDCNQWLSDWVVKSEYKEPGIARCAAGEMADLLLTTPSKKFTCQGPV 907
Db 850 SHLAGINPLYCDCSLRWLSEWIKAGYKEPGIARCSSPESMADRLTLTPHTRFQCKGPV 909
Qy 908 DVNILAKNCPCLSNPKNDGTCSNDPVDYFRTCTPYGFGQDCDVPFHACISNPKKHGGT 967
Db 910 DINIVAKNACLSSPCKNNGTCSQDPVEYRCTCPYSYKGDCTVPINTCVQNPQCHGGT 969
Qy 968 CHLKEGEEDGFWCICADGFEGENCEVNDDEDNDCENNSTCVGINNYTCLCPPEYTG 1027
Db 970 CHLSHSHRDGFSCSCPLGFGQRCINPDCCEDNDCENSATCVGINNYACVPPNYTG 1029
Qy 1028 LCEEKLDFAQDLNCPQHSKILTPKGFCDCTPGYVGEHCDIDFDCQDNKCKNGAHC 1087
Db 1030 LCDEVIDYCVPEMNLQHEAKCISLDKGRCEFCVPGYSGLKCETDNDCAHCRHGAQC 1089
Qy 1088 TDAVNGYTICPEGYGLFCEFSPPMVLPRTPCNDPDCNQAGCIVRINEPICQCLPGY 1147
Db 1090 VDAVNGYTICPQGSGLFCEHPPMVLQTPCNDPDCNQAGCIVVQEPTRCPCPGF 1149


```

DR   Dev. Biol. 0:0-0(1999).
DR   EMBL; AF144629; AAD44760.1; -.
DR   HSSP; P01132; 1EGF.
DR   MGD; MGI:1315202; Slit3.
DR   INTERPRO; IPR000152; -.
DR   INTERPRO; IPR000359; -.
DR   INTERPRO; IPR000372; -.
DR   INTERPRO; IPR000483; -.
DR   INTERPRO; IPR000561; -.
DR   INTERPRO; IPR000742; -.
DR   INTERPRO; IPR001010; -.
DR   INTERPRO; IPR001438; -.
DR   INTERPRO; IPR001611; -.
DR   INTERPRO; IPR001791; -.
DR   INTERPRO; IPR001881; -.
DR   INTERPRO; IPR002049; -.
DR   INTERPRO; IPR002272; -.
DR   INTERPRO; IPR002400; -.
DR   PFAM; PF000008; EGF; 9.
DR   PFAM; PF00054; laminin_G; 1.
DR   PFAM; PF00560; LRR; 19.
DR   PFAM; PF01462; LRRNT; 4.
DR   PFAM; PF01463; LRRCT; 4.
DR   PRINTS; PR00010; EGFBL00D.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00019; LEURICHRPT.
DR   PRINTS; PR00287; THIONIN.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   PRINTS; PR01143; FSHRECEPTOR.
DR   PROSITE; PS000010; ASX_HYDROXYL; UNKNOWN_2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR   PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 7.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
KW   Glycoprotein; EGF-like domain.
FT   NON_TER   1523  1523
SO   SEQUENCE   1523 AA: 167710 MW:  F43A3F3E016C4BFC CRC64:

```

Query Match 68.6%; Score 5703; DB 11; Length 1523;
Best Local Similarity 66.8%; Pred. No. 0;
Matches 1014; Conservative 222; Mismatches 267; Indels 14; Gaps 9;

```
Qy      11 LSLGLVLA-ILNKVAPQAACPAQCSCSGSTVDCHGLALRSVPNIPRTERLDLNGNNITR   69  
       |:| | | :| | :| | :| | | | | | | | | | | | | | | |  
Db      16 LALGLALASISLGPPAAACPCTKCTCSAASVDCHGLGLRAVRPGIPRNAERLDLDRNNITR   75  
  
Qy     70 ITKTDFAGLRHLRVQLMENKISTIERGAFQDLKELERLRNRNHQLPPELLFLGTAKL   129  
       || | | | | | :| | | | | :| | | | | | | | | | | | | | | |  
Db     76 ITKMDFAGLNKLRLVHLEDNQVSIIERGAFQDLKQLERLRNLKNKLQVLPELLFQSTPKL   135  
  
Qy    130 YRLDLSENQIAIPRKAFRGAVDIKNQLDYNIQISCIEDGAFRALRDLEVLTLNNNNITR   189  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    136 TRLDLSENQIQIGIPRKAFRGVTGVKNQLDNNHISCIEDGAFRALRDLEITLTLNNNNISR   195  
  
Qy    190 LSVASFNMHPKLRITFRLHSNNLYCDCHLAWLSDWLKRPRVGLYTCMGPSHLRGHNVAE   249  
       : | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    196 ILVTSFNHMPIRTLRHSNNHYCDCHLAWLSDWLQRRTIQQFTLCMAPVHLRGFSVAD   255  
  
Qy    250 VQKREFVCSDEEEGHQSFMAPSCSV--LHCPAACTCSNNIVDCRGKGLTEIPTNLPETIT   307  
       || | | :| | | | | | | | | | | | | | | | | | | | | |  
Db    256 VQKEKYCV---PGPHS-EAPACNANSLSCPACSCTSNNVDCRGKGLTEIPANPEGIV   310  
  
Qy    308 EIRLEQNTIKVIPPGAFFSPYKKLRIDLNNQISELPADFQGLRSLNSLVLYGNKITEL   367  
       || | | | | :| | | | | | | | | | | | | | | | | | | | | |  
Db    311 EIRLEQNSIKSIPAGAFTQYKKLRIDISKQISDIAPDAFQGLKSLSLVLYGNKITEI   370  
  
Qy    368 PRKSLFEGLFSQLLLLLNANKINCLRVDAFDQLHNLNLLSYDNKLTQIAKGTFSPRLAIQ   427  
       || | | :| | | | | | | | | | | | | | | | | | | | | |  
Db    371 PKGLFDGLYSQLLLLLNANKINCLRVTFDQLNLNLLSYDNKLTQISKGLFVPLQSQI   430  
  
Qy    428 TWHLAONPFIDCHLKWLAELYHTNPETISGARCTSPRRLANKRIGQIKSKKFCRSGTED   487
```

```

RESULT 7
Q9WVB4
ID Q9WVB4 PRELIMINARY; PRT; 1523 AA.
AC Q9WVB4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT3 (FRAGMENT).
GN SLIT3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/ICR;
RA Yuan W., Zhou L., Chen J.-H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse SLIT family: Secreted ligands for Robo expressed in
RT patterns that suggest a role in morphogenesis and axon guidance.";
```

Db 431 TLHLAQNPFVCDCHLKLADVLQDNPIETSGARCSSPRRLANKRISQIKSKKFRCSGSED 490

Qy 488 YRSKLSGDCFADLACPEKCRCEGTTVDCSNQKLNKPIEHIPOYTAEELRNNNEFTVLEAT 547
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 491 YRNRFSSECFMDLVCPKCRCEGTTIVDCSNQKLARIPSHLPEYTTDLRLNDNDISVLEAT 550

Qy 548 GIFKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLNVOHKMFKGLESKLT 607
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 551 GIFKLPQLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLNVOHKMFKGLESKLT 610

Qy 608 MLRSNRITCVGNDSPFGLSSVRLSLYDQNTTAVGAFDTHLSLSTLNLLANFPNCNCY 667
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 611 MLRSNLISCVSNDTFAGLSSVRLSLYDQNTTAVGAFDTHLSLSTLNLLANFPNCNCY 670

Qy 668 LAWLGEWLKRRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNDSCLSPRCPTC 727
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | |

Db 671 MAWLGEWLKRRIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDGNDNDSCLSPRCPTC 729

Qy 728 TCDLTVVRCNKLKVLKPGIPRDVTELYLDGNQFTLVKELSNYKHLTLIDLNNRIST 787
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 730 TCVETVRCNKLKVLKPGIPRDVTELYLDGNQFTLVKELSNYKHLTLIDLNNRIST 789

Qy 788 LSNQSFNMTOQLTLILSYNRLRCIPPTFDGLKSLRLSLHNDISVVEGAFNDLSAL 847
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 790 LTNHTFSMNSHSTLILSYNRLRCIPPTFDGLKSLRLSLHNDISVVEGAFNDLSAL 849

Qy 848 SHLAIGANPLYCDNMQLSDWVKSEYKEPGIARCAPGEMADKLLTTPSKKTCQGPV 907
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 850 SHLAIGANPLYCDNMQLSDWVKSEYKEPGIARCAPGEMADKLLTTPSKKTCQGPV 909

Qy 908 DVNLLAKNCPCLSPCKNDGTCNSDPVDFYRCTCPYFGKQDCDPIHACISNPKHGGT 967
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 910 DINIVAKNCPCLSPCKNDGTCNSDPVDFYRCTCPYFGKQDCDPIHACISNPKHGGT 969

Qy 968 CHLKEGEDGFWICADGFEENCEVNVDDCENDCENNSTVDGINNYTCLPPEYTG 1027
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 970 CHLKEGEDGFWICADGFEENCEVNVDDCENDCENNSTVDGINNYTCLPPEYTG 1029

Qy 1028 LCEEKLDFAQDLNCPQHDSKILTPKFGKDCDTPYGVYGEHCDIFDDCQDNKNGAHC 1087
| | | : | : | : | : | : | : | : | : | : | : | : | : |

Db 1030 LCDEVIDYCPVEMNLQHEAKISLDKGFRCBVPYSGKLCETNNDCAVAKRCHGAQC 1089

Qy 1088 TDAVNGYTCICPEGYSGLCFEPPMVLPTSPCDNFCQNGAQCIVRINEPICQLPGY 1147
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 1090 VDEVNGYTCICPEGYSGLCFEPPMVLPTSPCDNFCQNGAQCIVRINEPICQLPGY 1149

Qy 1148 QGEKCEKLVSNFINKESYLQIPSAKVRPQNTITLQIATDEDSGILLYKGDNDHIAVELY 1207
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | |

Db 1150 AGPRCEKLITVNFVGKDSYVELASAKVRPQANISLQVATDKDNGILLYKGDNDHIAVELY 1209

Qy 1208 RGRVRSYDTGSHPASAIYSVETINDGNFHVILLALDQSLSLSDVGGNPKIITNLSKQS 1267
: | | | : | | | : | | | : | | | : | | | : | | | : | | |

Db 1210 QGHVRLVYDLSPPPTTVSVETVNDGQFHSVKLVMLNQTNLNVVDKGAPKSLGKLOKQ 1269

Qy 1268 TLNFDSPLYVGGMPGKSNVASLRQAPGNGTSFHHGCIIRNLYINSELQDFQKVPMT-GIL 1326
: | | | : | | | : | | | : | | | : | | | : | | | : | | |

Db 1270 AVGSNPLYLGGIPTSTGLSALRQAGDRPLGGFHCHEVRINNELQDFKALPPQSLGVS 1329

Qy 1327 PGCEPHCKKVCAGHTCQPSQAGFTCECGEGWGPLCDQRTNDPCLGNKCVHGTCLPINA 1386
| | | : | | | : | | | : | | | : | | | : | | | : | | |

Db 1330 PGCKSC--TVCRHGLCRSEKDSVVCCEHPGWTGPLCDQEARDPCLGHSCRHGTGM-ATG 1386

Qy 1387 FSYSCCKLEHGGVLCDEEEDLFNCPQAIKCKHGCRLSGLGQPYCESSGYTGDSCDRE 1446
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 1387 DSYVCKCAEYGGALCDQKNDASASAFKCHHGCCHISDRGEPCYLCQPGFSGHHCBE 1446

Qy 1447 ISCRGERIRDYQKQGYAACQTTKVRSLRECRGGAGGCCPLRRKRYKSFECTDGS 1506
| | | : | | | : | | | : | | | : | | | : | | | : | | |

Db 1447 NPCMGEIVRAIRKQDYASCATASKVPIMECRGCC-GSQCCQPIRSKRRKIVFQCTDGS 1505

Qy 1507 SFVDEVEKVKCGCTRC 1523
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 1506 SFVDEVEKVKCGCTRC 1522

RESULT 8

075093

ID 075093 PRELIMINARY; PRT; 1534 AA.
AC 075093;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT-1 PROTEIN.
GN SLIT-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99033071; PubMed=9813312;
RA Itoh A., Miyabayashi T., Ohno M., Sakano S.;
RT "Cloning and expressions of three mammalian homologues of Drosophila
RT slit suggest possible roles for Slit in the formation and maintenance
RT of the nervous system."
RL Brain Res. Mol. Brain Res. 62:175-186(1998).
DR EMBL; AB017167; BAA35184.1; -.
DR HSSP; P00743; ICCF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 19.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGFAMININ.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01185; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 2.
KW Glycoprotein; EGF-like domain.
SQ SEQUENCE 1534 AA; 167951 MW; 8954EF8EA4DAEBA1 CRC64;

Query Match 67.3%; Score 5597; DB 4; Length 1534;
Best Local Similarity 65.3%; Pred. No. 0;
Matches 992; Conservative 228; Mismatches 283; Indels 16; Gaps 5;

Qy 15 LVLAILNKVAPQACPAQCSGSGSTVDCHGLALRSVPRNIPRTERLDLNGNNITRITKTD 74
| : | : | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 21 LWWAAWRLGASACPALCTCTGTVDCHGTGLQAIKPNIPRTERLELNGNNITRIKND 80

Qy 75 FAGLRLRLVQLMENKISTIERGAFQDLKELERLRLNRHLQLPPELLFLGTAKLYRLDL 134
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 81 FAGLRLRLVQLMENKISTIERGAFQDLKELERLRLNRHLQLPPELLFLGTAKLYRLDL 140

Qy 135 SENQIAIPKRAFRGAVDIKQLDYNQISCIEDGAFRALRDLEVLTLNNGNITRLSVAS 194
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 141 SENQIAIPKRAFRGATDLKRLDKNQISCIIEGAFRALRGLEVLTLNNGNITRIPVSS 200

Qy 195 FNMHPKLRTRFLSHNNLYCDCHLAWLSDWLRKRPRVGLTYCQMGPSHLRGNVAVQKRE 254
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 201 FNMHPKLRTRFLSHNNLYCDCHLAWLSDWLRKRPRVGLTYCQMGPSHLRGNVAVQKSE 260

```

RESULT 9
O88279
ID O88279 PRELIMINARY; PRT; 1531 AA.
AC O88279;
DT 01-NOV-1998 (TrEMBLRel. 08, Created)
DT 01-NOV-1998 (TrEMBLRel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLRel. 15, Last annotation update)
DE MEGF4.
GN MEGF4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RL like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
DR EMBL; AB011530; BAA32460.1; -.
DR HSSP; P00743; 1CCF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 19.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PRO0010; EGFBLD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
SO SEQUENCE 1531 AA: 167497 MW: DFC4B60CCBC5529A CRC64:

```

Qy 15 LVLAILNKVAPQACPAQCSCSGSTVDCHGLALRSVPNIPNTERLDLNGNNITRITKTD 74
| : | : : ||| |:|:||||| |:|:|||||||:||||||| | :|
Db 21 LLWAAANRIGATACPACTCTGTTVDCHGTGLQAIKPNIPNTERLELNGNNITRIHKND 80

[illegible]

09WVB5

```

ID Q9WVB5 PRELIMINARY; PRT; 1531 AA.
AC Q9WVB5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT1.
GN SLIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/ICR;
RA Yuan W., Zhou L., Chen J.-H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse SLIT family: Secreted ligands for Robo expressed in
RL patterns that suggest a role in morphogenesis and axon guidance.";
RL Dev. Biol. 0:0-0(1999).
DR EMBL; AF144627; AAD44758.1; .
DR HSSP; P00743; 1CCF.
DR MGD; MGI:1315203; Slit1.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 19.
DR PFAM; PF01462; LRRNT; 4.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PRO00010; EGF_BLOOD.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 2.

```


DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF000008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 7.
DR PFAM; PF01462; LRRNT; 2.
DR PFAM; PF01463; LRRCT; 2.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00019; LEURICHRPT.
DR PRINTS; PR00287; THIONIN.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 2.
DR PROSITE; PS01225; CTCK_2; 1.
KW Glycoprotein; EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 1025 AA; 112974 MW; 46CD0D5B7246FC72 CRC64;

Query Match 66.5%; Score 5530; DB 11; Length 1025;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 983; Conservative 25; Mismatches 17; Indels 0; Gaps 0;

Qy 501 ACPEKRCCEGTTVDCSNQRLNKIPEHIPQYTAELRLNNEFTVLEATGIFKKLPQLRKIN 560
Db 1 ACPEKRCCEGTTVDCSNQRLNKIPDHIPQYTAELRLNNEFTVLEATGIFKKLPQLRKIN 60
Qy 561 FSNKKTIDIEGAFEGASGVNEILLTSNRLNVQHKMPGLSLKTLMLRSNRITCVGND 620
Db 61 FSNKKTIDIEGAFEGASGVNEILLTSNRLNVQHKMPGLSLKTLMLRSNRISCVGND 120
Qy 621 SFIGLSSVRLSLYDNIQITVAPGAFDTLHSLSTLNLNLANFNCNCLAWLGEWLKKRI 680
Db 121 SFIGLSSVRLSLYDNIQITVAPGAFDXLHSLSTLNLNLANFNCNCHLAWLGEWLKKRI 180
Qy 681 VTGNPRCQKPYFLKEIPQDVAIQDFTCDGNDNCSPLSRCPSECTCLDTVVRCSNKG 740
Db 181 VTGNPRCQKPYFLKEIPQDVAIQDFTCDGNDNCSPLSRCPSECTCLDTXVRCSNKG 240
Qy 741 LKVLPGKIPRDVTELYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQSFNMTOQL 800
Db 241 LKVLPGKIPKDVTELYLDGNQFTLVPKELSNYKHLTLIDLNNRISTLSNQXFSNMTOQL 300
Qy 801 TLILSYNRLRCIPPTFDGLSLRLSLHGNDISVVEGAFNDLSALSHLAIGANPLYCD 860
Db 301 TLILSYNRLRCIPPTFDGLSLRLSLHGNDISVVEGAFNDLSALSHLAIGANPLYCD 360
Qy 861 CNMQWLSDWKSEYKEPGIARCAAGPGEMADKLLLTTPSKKFTCQGPVDVNLAKNCPCLS 920
Db 361 CNMQWLSDWKSEYKEPGIARCAAGPGEMADKLLLTTPSKKFTCQGPMDITIQAACNCPCLS 420
Qy 921 NPCKNDGTCNSDPVDFYRCTCPYGFQKQDCDVPFHACISNPCKHGGTCHLKEGEEDGFWC 980
Db 421 NPCKNDGTCNSDPVDFYRCTCPYGFQKQDCDVPFHACISNPCKHGGTCHLKEGENAGFWC 480
Qy 981 ICADGFECEGNEVVDCEDENDCENNSTCVGINNYTCLCPPEYTGELCEKLDFAQDL 1040
Db 481 TCADGFECEGNEVVDCEDENDCENNSTCVGINNYTCLCPPEYTGELCEKLDFAQDL 540
Qy 1041 NPCQHDSKCLTPKGFCDCTPGYVGEHCDIDFDDQDNCKKNGAHTDAVNGYTICPE 1100
Db 541 NPCQHDSKCLTPKGFCDCTPGYVGEHCDIDFDDQDNCKKNGAHTDAVNGYTCVPE 600
Qy 1101 GYSGLFCEFSPPMVLPRTPSCDNFDCQGAQCIIVRINEPICQLPGYQGEKCEKLVSVNF 1160
Db 601 GYSGLFCEFSPPMVLPRTPSCDNFDCQGAQCIIVRINEPICQLPGYQGEKCEKLVSVNF 660

Qy 1161 INKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRSYDTGSH 1220
Db 661 VNKESYLQIPSAKVRPQTNITLQIATDEDSGILLYKGDKDHIAVELYRGRVRSYDTGSH 720
Qy 1221 PASAIYSVETINDGNFHVIELLDQSLSLSDGNGNPKIITNLSKQSTLNFDSPLYVGGM 1280
Db 721 PASAIYSVETINDGNFHVIELLDQSLSLSDGNGSPKIVITNLSKQSTLNFDSPLYVGGM 780
Qy 1281 PGKSNVASLRQAPGQNGTSFHGCIIRNLYINSELQDFQKVPMTGTILPGCEPCHKKVCAHG 1340
Db 781 PGKNNVASLRQAPGQNGTSFHGCIIRNLYINSELQDFRKMPTGTILPGCEPCHKKVCAHG 840
Qy 1341 TCQPSSQAGPTCECEGEGWMLPLCDQRTNDPCLGNKCVHGTCLPINFASYSCCKLEHGGV 1400
Db 841 MCQPSSQSGPTCECEGEGWMLPLCDQRTNDPCLGNKCVHGTCLPINFASYSCCKLEHGGV 900
Qy 1401 LCDEEDLFPNCPQAIKCKHGKRLSGLPQYCECSSGYTGSDREISCRGERIRDYQK 1460
Db 901 LCDEEDLFPNCPQAIKCKHGKRLSGVGPYCECNSGFTGSDREISCRGERIRDYQK 960
Qy 1461 QQGYAACQTTKKVSRLECRGGCAGGCGCCPLSRKRRYSFECTDGSFVDEVEKVVKCGC 1520
Db 961 QQGYAACQTTKKVSRLECRGGCAGGCGCCPLSRKRRYSFECTDGSFVDEVEKVVKCGC 1020
Qy 1521 TRCVS 1525
Db 1021 ARCAS 1025

RESULT 12
Q9WUG5
ID Q9WUG5 PRELIMINARY; PRT; 1530 AA.
AC Q9WUG5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
DR EMBL; AF133730; AAD25540.1; -.
DR HSSP; P00743; 1CCF.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF000008; EGF; 9.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF00560; LRR; 18.
DR PFAM; PF01462; LRRNT; 3.
DR PFAM; PF01463; LRRCT; 4.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00019; LEURICHRPT.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
DR PROSITE; PS01185; CTCK_1; UNKNOWN_1.

DR PRINTS: PR00019: LETHICRPT

Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196005; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.C.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003809; AAF58097.1; -.
 DR HSSP: P00743; 1CCF.
 DR FLYBASE; FBgn0003425; sli.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000359; -.
 DR INTERPRO: IPR000372; -.
 DR INTERPRO: IPR000483; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR000742; -.
 DR INTERPRO: IPR001611; -.
 DR INTERPRO: IPR001791; -.
 DR INTERPRO: IPR001881; -.
 DR INTERPRO: IPR002049; -.
 DR PFAM: PF00007; Cys_knot; 1.
 DR PFAM: PF00008; EGF; 7.
 DR PFAM: PF00054; laminin_G; 1.
 DR PFAM: PF00560; LRR; 17.
 DR PFAM: PF01462; LRRNT; 4.
 DR PFAM: PF01463; LRRCT; 4.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR00019; LEURICHRPT.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR PROSITE; PS01225; CTCK_2; 1.
 SO SEQUENCE 1504 AA; 168597 MW; 836A3F5022BF234F CRC64;

```
Qy      1044 QHSDSKCILTPKGFCDCTPGYVGEHCDIDFDDCQDNCKNGAHACTDAVNGYTCICPEGYS 1103  
        : :| | : :|| : |:| : |||:: :|| | :| :| :| :|  
Db      1060 ANGAKCMDFHTHYSCDCQAGFHGTNCTNDIDDQNHCMQNGGTCVDGINDYQCRCPPDYT 1119  
        : :| | : :||| : :|| : | : : :| | | | :| :| :| :|  
Qy      1104 GLFCFESP--PMVLPRTPSPCNDFDCQNGA--QCIVRINEPICQLPGYQGEKEKLSVSN 1159  
        | :| | :| :| | | : :|| : | : : :| | | | :| :| :| :|  
Db      1120 GKYCEGHNMISMYPQTSPQCNHECHKGVCFQPNAQGS DYLCRCHPGYTGNWCEYLT SIS 1179  
        : :| | : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1160 FINKESYLQIPSAKVPRPTNTILQIATDEDSGLLYKGDKDHIAVELYGRVRASYDTGS 1219  
        | :| : : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1180 FVHNNSFVELPLRLTRPEANVTIFSSAEQNGILMYDGQDAHLAVELFNIRVSYDVGN 1239  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1220 HPASAIYSVETINDGNFIHVALLDQSLSLSDVGNGPKIIITNLKSQSTLNFDSPLYVGG 1279  
        | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1240 HPVSTMYSFEMVADGKYHAVELLAIKKNFTLRVDRLARSINEGSNDYLLKLTTPMFLGG 1299  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1280 MPGKSNVASLRQAQGNGTSFGHCIRNLYINSELQDFVKVPMTQGITLPCEPCHKVKCAH 1339  
        : | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1300 LPVDPAQQAYKNWQIRNLTSFGKMKEVWINHKLVDFGNAQRQOKITPGC-----ALLE 1353  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1340 GTCQPSSQAGTFCECQEQWMG--PLCDQRTNDPLCGNKCVHGT-CLP-INA-FSYSCKCL 1394  
        | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1354 GEQQEEB-----DDEQDFMDTPHIKEEPVDPCLENKCRGRSGCPVSNARNRGYQCKCK 1407  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1395 EHGHSVLCDEEDLFNPQQAICKKHGKRSLGLGPYCECSSGYTGDSCDREISCRGERI 1454  
        | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1408 HQQGRYICDQEGESTEP-----PTVTAAS-----TCRKEQV 1438  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Qy      1455 RDYYQKQGYAQACOTTKKVSRLCEKGGCAGGCCGGLPSRKRRKYSPECTDSSSFVDEVK 1514  
        | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db      1439 REYTTEND---CSRQPLYAKCVGC-GNQCCA KIVRRKKVRMVCSNNRKYIKNLDI 1493  
        : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
  
Qy      1515 VVKCGCTR 1522  
        | | | | | :  
Db      1494 VRKCGCTK 1501
```

```

ID Q9XYV4 PRELIMINARY; PRT; 1504 AA.
AC Q9XYV4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SLIT PROTEIN.
GN SLIT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200390; PubMed=10102267;
RA Kidd T., Bland K.S., Goodman C.S.;
RT "Slit is the midline repellent for the robo receptor in Drosophila.";
RL Cell 96:785-794(1999).
DR EMBL; AF126540; AAD26567.1; -.
DR HSSP; P00740; 1EDM.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000359; -.
DR INTERPRO; IPR000372; -.
DR INTERPRO; IPR000483; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001611; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR INTERPRO; IPR002049; -.
DR PFAM; PF00007; Cys_knot; 1.
DR PFAM; PF00008; EGF; 7.
DR PFAM; PF00054; laminin_G; 1.
DR PFAM; PF01462; LRRNT; 4.

```

Search completed: January 22, 2001, 12:48:00
Job time: 1681 sec :

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:15:50 ; Search time 233.01 Seconds
(without alignments)
204.714 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPMPENHAIARSTSTNN.....SCLYAEAGEPAPQMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	7427	100.0	1395	20	Y13563	Drosophila Robo 1
2	7427	100.0	1395	20	Y08401	Drosophila sp. ROB
3	1788	24.1	1380	20	Y08402	Drosophila sp. ROB
4	1786.5	24.1	1381	20	Y13564	Drosophila Robo 2
5	1592	21.4	1651	20	Y13566	Human Robo 1 polyp
6	1588	21.4	1297	20	Y13565	C. elegans Robo po
7	1588	21.4	1297	20	Y08403	C. elegans ROBO pr
8	1584	21.3	1649	20	Y08404	Human ROBO1 protei
9	1317.5	17.7	753	20	W83927	Human T85 protein.
10	665.5	9.0	1571	19	W42087	Human Down syndrom
11	661	8.9	1910	19	W42086	Human Down syndrom
12	639	8.6	1728	12	R13144	Deleted in Colorec

13	637.5	8.6	1447	16	R68553	Deleted in colorec
14	637.5	8.6	1447	20	Y33498	Human DCC protein.
15	633.5	8.5	1257	20	W74152	Human L1 cell adhe
16	614.5	8.3	1018	15	R63759	Human contactin (F
17	614.5	8.3	1018	17	R87028	Human contactin.
18	610.5	8.2	1192	19	W57900	Protein of clonin C
19	603	8.1	1299	21	Y40439	Human Nr-CAM prote
20	602	8.1	1028	19	W29667	Homo sapiens DL185
21	588	7.9	1018	18	W06485	Rat contactin like
22	586.5	7.9	1304	19	W59994	Human neurai cell
23	579.5	7.8	1911	16	R71726	Human PTP-OB. Hom
24	579.5	7.8	1911	18	W27225	Human protein tyro
25	579.5	7.8	1911	20	W94027	Human protein tyro
26	573	7.7	1496	20	W81030	Melanoma associat
27	573	7.7	1496	21	Y70469	Human p53 target m
28	548	7.4	4412	21	Y53666	Sequence gi/101742
29	546	7.4	1897	21	Y81785	Human protein tyro
30	546	7.4	1897	21	Y56100	LAR tyrosine phosph
31	545	7.3	434	20	Y13567	Human Robo 2 polyp
32	545	7.3	434	20	Y08405	Human partial ROBO
33	529.5	7.1	1501	16	R72858	Rat receptor type-
34	526	7.1	1242	19	W52287	Rattus norvegicus
35	510	6.9	985	20	Y41716	Human PRO860 prote
36	509	6.9	1225	19	W52289	Homo sapiens cdo t
37	490	6.6	3117	21	Y53667	Sequence gi/332818
38	480.5	6.5	1070	18	W08747	Human colon carcin
39	458	6.2	1853	21	Y53668	Protein 608 sequen
40	457	6.2	2387	21	Y53665	Mechanical stress
41	457	6.2	2597	21	Y53664	Mechanical stress
42	446	6.0	1139	19	W37779	Rattus norvegicus
43	440.5	5.9	1125	19	W52288	Rattus norvegicus
44	439	5.9	848	21	Y88565	Human NCAM 140kd 1
45	438.5	5.9	1251	19	W37778	Rattus norvegicus

ALIGNMENTS

RESULT 1
Y13563
ID Y13563 standard; Protein; 1395 AA.
XX
AC Y13563;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55767.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 30-33; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 1395 AA;

Query Match 100.0%; Score 7427; DB 20; Length 1395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHPMPHENAIARSTSTNNPSRGRSSRMWLLPAWLLLVASNGLPAVRGQYQSPRIE 60
Db 1 mhpmpphenaiarststnnpsrgrssrmwllpawlllvvasnglpavrgqyqsprie 60
Qy 61 HPTDLVVKNEPATLNCKVEGKPEPTIEWFKDGEVPSTNEKSHRVQFKDGALFFYRTMQ 120
Db 61 hptdlvvknepatlnckvegkpeptiewfkdgevpstnekkshrvgfkdgalfyrtmq 120
Qy 121 GKKEQDGGGEYWCVAKNRVGQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Db 121 gkkeqdgggeywcvaaknrvvgqavsrhaslqiavlrddfrvepkdtrvakgetallecgppk 180
Qy 181 GIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNNLLISNVEPIDEGNYKCIAQNLVG 240
Db 181 gipeptliwikdgvplddlksamsgassrvrivdggnllisnvepidegnykciaqnlvg 240
Qy 241 TRESSYAKLIVQVKPYFMKEPKDQVLMYGGQTATFHCSVGGDPPPKVLWKEEGNIPVSR 300
Db 241 tressyaklivqvkpyfmkepkdqvmlmyggqtatfhcsvvggpppkvlwkeegnipvsr 300
Qy 301 RILHDEKSLEISNITPTDEGTIVCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVGLNG 360
Db 301 rilhdeksleisnitptdegtyvceahnnvgqisaraslivhappnftkrpsnkkvglng 360
Qy 361 VVQLPCMASGNPPPSVFWTKEGVSTLMPFNSSHGRTQYVAADGTLQITDVRQDEGEYVCS 420
Db 361 vvqlpcmasgnpppsvfwtkegvstlmpfnsshgtrqyvaadgtlqitdvrqdegeyyvcs 420
Qy 421 AFSVDSSTVRVQLQVSSVDERPPPIIIGPANQLPKGVSATLPCRATGNPSRIKWFH 480
Db 421 afsvdsstvrvlqvvssvderpppiigpanqlpkgvsatlpcratgnpsrikwfh 480
Qy 481 DGHAVQAGNRYSIIGSSLRVDDLQLSDSGTYTCTASGERGETSWAATLVEKPGSTSLH 540
Db 481 dghavqagnrysiigsslrddllqldsgstyctasgergetswaatlvekpgstslh 540
Qy 541 RAADPSTYPAPPGTTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGYTVYFSPDLQTGWIV 600
Db 541 raadpstypappgttpkvlnvstrtsislrwaksqekpgavgpiigytyvfspdlqtgwiv 600
Qy 601 AAHRVGDQTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSA 660
Db 601 aahrvgdtqvtisgltpgtsyvlvraentqgisvpsglsnvktieadfdaasandlsa 660
Qy 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDSVPSAQYHSITV 720
Db 661 artlltgksvelidasainasavrlewmhvsadekeyveglrihykdsavpsaqyhsitv 720
Qy 721 MDASAEFVVGNLKKYTKYEFFLTPFFETIEGQPSNSKTALTYEDVPSAPPDNIQIGMTN 780
Db 721 mdasaefvvgnlkkytkyeffltpffetiegqpsnsktaltyedvpsappdniqigmtn 780
Qy 781 QTAGWVRWTPPPSQHHNGNLGYKIEVSAGNTMKVLNMTLNATTTSVLLNLTGAVIS 840
Db 781 qtagwvrwtpppsqqhngnlygykievsagntmkvlnmtlnatttsvllnltgavys 840

Qy 841 VRLNSFTKAGDGPYSKPSISLFMDPTHVHPPRAHPSGTHDGRHEGQDLTYHNNGNIPPGD 900
Db 841 vrlnsftkagdgypyskpislfmdpthvhpprahsptgdgrhegqdltyhnngnippgd 900
Qy 901 INPTTHKKTTDYLSGPWLMVLCIVLLVLVISAISMVYFKRKHQMTKELGHLSSVSDNE 960
Db 901 inptthkktttdylsgpwlmlvcivllvlvisaismvyfkrkhqmtkelghlsvsdne 960
Qy 961 ITALNINSKESLWIDHRGWRTADTKDGLSESKLLSHVNSSQSNYNNSDGGTDYAEVD 1020
Db 961 italninskeslwidhrgwrtadtkdglseksllshvnssqsnynnsdggtdyaevd 1020
Qy 1021 TRNLTFYNCRKSPDNPTPYATTMIIGTSSSETCTKTTISADKDSGTHSPYSDAFAGQV 1080
Db 1021 trnltfyncrkspdnptpyattmiigtsssetctkttisadkdsghthspysdafagqv 1080
Qy 1081 PAVPVVKNYLYQYVEPINWSEFLPPPEHPPPSTYGYAQGSPSSRSKSSKAGSGIST 1140
Db 1081 pavpvvksnylyqyvepinwseflpppehppsstygyaqgspssrsksskagsgist 1140
Qy 1141 NQSILNASIHSSSSGGFSAWGSPQYAVACPPENVSNPLSAVAGGTQNRQYITPTNQHP 1200
Db 1141 nqsilnasihssssggfsawgspqyavacppenvsnplsavaggtqnryitptnqhp 1200
Qy 1201 POLPAYFATTGPGGAVPNHLPFATQRHAASEYQAGLNAARCAQSRACNSDALATPSM 1260
Db 1201 polpayfattgpggavpnhlpfatqrhaaseyqaglnaarcaqsracnsdalatpsm 1260
Qy 1261 QPPPPVPVPEGWYQPVHPNSHPMPTSSNNHQIYQCSSECSHRSQSHKRLQLEEHS 1320
Db 1261 qppppvpvpegwyqpvhpnshpmptssnnhqiyqcssecsdhsrsgshkrqlleehs 1320
Qy 1321 SAKQRGGHRRRAPVVPQCMESENENMLAEYEQRYTSDCCNSSREGDTSCSEGSCLYA 1380
Db 1321 sakqrghhrrrapvvqpcmesenenmlaeyeqrytsdcccnsregdtscsegsclya 1380
Qy 1381 EAGEPAPRQMTAKNT 1395
Db 1381 eagepaprmtaknt 1395

RESULT 2

Y08401

ID Y08401 standard; Protein; 1395 AA.

XX

AC Y08401;

XX

DT 24-JUL-1999 (first entry)

XX

DE Drosophila sp. ROBO1 protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX

OS Drosophila sp.

XX

PN W09920764-A1.

XX

PD 29-APR-1999.

XX

PF 20-OCT-1998; 98WO-US22164.

XX

PR 14-NOV-1997; 97US-0971172.

PR

20-OCT-1997; 97US-0062921.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman CS, Kidd T, Mitchell KJ, Tear G;

XX

DR WPI; 1999-312615/26.

DR

N-PSDB; X57250.

XX

PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 45-49; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1395 AA;

Query Match 100.0%; Score 7427; DB 20; Length 1395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MHPMHPENHAIARSTSTNNPSRSRSMWLLPAWLLVLVSNGLPAVRGYQSPRIIE 60
Db 1 mhpmpenhaiarststnnpsrsmwllpawllvlvsnvglpavrgyqsprie 60
Qy 61 HPTDLVVKNEPATLNCKVEGKPEPTIEWFKDGEVSTNEKSHRVQFKDGALFFYRTMQ 120
Db 61 hptdlvvknepatlnckvegkpeptiewfkdgepstnekskhrvqfkdgalffyrmtq 120
Qy 121 GKKEQDGGYWCVAKNRVGQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Db 121 gkkekdggywcvaknrvqgavsrhaslqlavlrddfrvepkdtrvakgetallecgppk 180
Qy 181 GIPEPTLIWIKDGVPLDDLKAMSFAGSSRVRLVDGNNLLISNVEPIDEGNYKICIAQNLVG 240
Db 181 gipeptliwikdgvplddlkamsfgassrvrlvdggnllisnvepidegnykiciaqnlvg 240
Qy 241 TRESSYAKLIVQVKPYFMKEPKDQVMYLGQTATFHCSVGGDPPPKVLWKEEGNIPVSRA 300
Db 241 tressyaklivqvkpyfmkepkdqvmlygqtatfhcsvvgdpppkvlwkeegnipvra 300
Qy 301 RILHDEKSLEISNITPTDEGTIVCEAHNNVGGISARASLIVHAPPNFKRPSNKKVGLNG 360
Db 301 rilhdeksleisnitptdegtyvceahnnvggisaraslivhappnftkrpsnkkvglng 360
Qy 361 VVQLPCMASGNPPSVFMTKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQEDGYVCS 420
Db 361 vvqlpcmasgnppsvfwtkegvstlmfpnsshgrqyvaadgtlqitdvrqedegyyvcs 420
Qy 421 AFSVDSSTVRVFLQVSSVDERPPPIIIGQANQTLPGKSVATLPCRATGNPSRIKWFH 480
Db 421 afsvdsstvrflqvssvderpppiigqanqtlpgksvatlpcratgnpsrikwfh 480
Qy 481 DGHAVQAGNRYSIQGSRLVDDLQSLDSGTYYTCTASGERGETSWAATLTVEKPGSTSLH 540
Db 481 dghavqagnrysiqgssrlvddlqslsdgttytctasgergetswaatltvekpgstslh 540
Qy 541 RAADPSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGYTVVEYFSPDLQGTGIV 600
Db 541 raadpstypapgtpkvlnvstsislrwaksqekpgavgpiigytyveyfspdlqgtgiv 600
Qy 601 AAHRVGDVTQVTSISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIADFDAAANDLSA 660
Db 601 aahrvgdvtqvtsisgltpgtsyvflvraentqgisvpsglsnviktiaedfdaandlsa 660
Qy 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKASVPSAQYHSITV 720
Db 661 artlltgksvelidasainasavrlewmhvsadekeyvegrihykdasvpsaqyhsitv 720
Qy 721 MDASAEFVVGNLKKYTKYEFFLTFFETIEGQPSNSKALTIEDVPSAPPDNIQIGMYN 780
Db 721 mdasaefvvgnlkkytkyeffltffetiegqpsnskaltiedvpsappdniqigmyn 780
Qy 781 QTAGVWRWTPPPSQHNGNLYGYKIEVSAGNTMKVLANMTLNATTSVLLNNLTGAVYS 840
Db 781 qtagvwrwtpppsghngnlygykievsagntmkvlanmtlnattsvllnnltgavys 840

Qy 841 VRLNSFTKAGDGPYSKPISLFMDPTHVHPPRAHPSGTHDGRHEGQDLTYHNNGNIPGD 900
Db 841 vrlnsftkagdgpysskpislfmdpthvhpprahsptgdgrhegqdltyhnngnipgd 900
Qy 901 INPTTHKKTDDYLSGFWMLVLCVILLVLVISAISMVYFKRKHQMTKELGHLVSVSDNE 960
Db 901 inptthkktddylsgfwmlvvcvillvlvisaismvfykrkhqmtkelghlsvsdne 960
Qy 961 ITALNINSKESLWIDHHRGWRTADTKDGLSESKLLSHVNSSQSNYNSDGGTDYAEVD 1020
Db 961 italninskeslwidhhrgrwtadtkdglseksllshvnssqsnynnsdggtdyaevd 1020
Qy 1021 TRNLTTFFYNCRKSPDNTPYATTMIIGTSSSETCTKTTISADKDSGTHSPYSDAFAGQV 1080
Db 1021 trnlttffyncrkspdnptpyattmiigtsssetctkttisadkdsghthspysdafagqv 1080
Qy 1081 PAVPVVKSNIQYVPEPINWSEFLPPPEHPSPSTYGYAQGSPESSRSKKSAGSGIST 1140
Db 1081 pavpvvksniqyvpepinwseflpppehpsstyggyaqgspeessrskksagsgist 1140
Qy 1141 NQSILNASIHSSSSGGFSANGVSPQYAVACPPENVYSNPLSAVAGGTQNRQYQITPTNHP 1200
Db 1141 nqsilnashssssggfsangvspqyavacppenvysnplsavaggtqnryqitptnqh 1200
Qy 1201 POLPAYFATTGPGGAVPPNHLFFATQRHAASEYQAGLNAARCAQSRACNSCDALATPSM 1260
Db 1201 polpayfattpgggavppnhlffatqrhaaseyqaglnaarcaqsracnscdalatpsm 1260
Qy 1261 QPPPPVPVPEGWYQVHPNHPMPTSSNHQIYQCSSECSDHSSSSQSHKRLQLEEHGS 1320
Db 1261 qppppvpvpegwyqvhpnhpmptssnhqiycssecdshsssqshkrlqleehgs 1320
Qy 1321 SAKQGGHRRRAPVVPQCMESNENMLAEYEQRYTSDCCNSREGDTCSCEGSCLYA 1380
Db 1321 sakqggghrrrapvvqpccmesenenmlaeyeqrytsdcccnsregdtcscegsclya 1380
Qy 1381 EAGEPAPRQMTAKNT 1395
Db 1381 eagepaprmtaknt 1395

RESULT 3
Y08402
ID Y08402 standard; Protein; 1380 AA.
XX
AC Y08402;
XX
DT 24-JUL-1999 (first entry)
XX
DE *Drosophila* sp. ROBO2 extracellular domain protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS *Drosophila* sp.
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57251.
XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G:

Qy 692 SADEKYVEGLRIHYKDASVP----- 711
: ||||| : : : |
Db 649 -ickvveafuvvarglppniynnnpyvtgntnllgsteteeasagacalietkopia 707

RESULT 5
Y13566

```

RESULT      5
Y13566
ID      Y13566 standard; Protein; 1651 AA.
XX
AC      Y13566;
XX
DT      30-JUL-1999 (first entry)
XX
DE      Human Robo.1 polypeptide.
XX
KW      Comm polypeptide; Robo polypeptide; commissureless; roundabout;
modulation; nerve cell function.
XX
OS      Homo sapiens.
XX
PN      WO9925833-A1.
XX
PD      27-MAY-1999.
XX

```

PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commisureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
XX function.
SQ Sequence 1651 AA;

Query Match 21.4%; Score 1592; DB 20; Length 1651;
Best Local Similarity 30.2%; Pred. No. 3.1e-83;
Matches 419; Conservative 188; Mismatches 492; Indels 290; Gaps 39;

Qy 56 PRIIEHPTDLVKKNEPATLNCKVEGKPEPTIEWFKDGEPVST--NEKSHRVQFKDGL 113
Db 68 privehpsdlivskgepatlnckaeqrptptiewykggervetdkddprshmlpsgsl 127
Qy 114 FFYRTMQGKKEQ-DGGEYWCVAKNRVQAVSRHASLQIAVLRRDDFRVEPKDTRVAKGETA 172
Db 128 fflrihvgrksrpdgeyyvvcarnylgeavshnaslevailrddfrqnpdvmvavgepa 187
Qy 173 LLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFGASSRVIVDGNLLISNVEPIDEGNYK 232
Db 188 vmecqpprghpeptiswkkdgspldd-----kderiti-rggklmitytrksdagkyv 239
Qy 233 CIAQLVGTRESSYAKLIVQVKPFMKPEKDVMLYQGATATFHCSVGGDPPPKVLWKKEE 292
Db 240 cvgtmvmgeresevaeltvlerpsfvkrpnsjvavtvdasafkceargdpvptvrwkdd 299
Qy 293 GNIPVSRARILHDEKSLEISNITPTDEGTIVCEAHNVGQISARASLIVHAPPNFTKRPS 352
Db 300 gelpksryei-rddhtlkirkvtagdmgsytcaenvmgkaeasatlvtqepghfvvkpr 358
Qy 353 NKKVGLNGVVQLPCMASGNPPSVFWTKEGVSTLMF---PNSSHGRQYVAADGTLQITDV 409
Db 359 dqvvalgrtvtfqceatgnppaifwrrregsqnlfsyqppqssrsfsvsqtgdltitnv 418
Qy 410 RQDEGGYVCSAFSVVDSSTVRVFLQVSSV-DERPPPIIQIGPANQTLPGKSVATLPCRA 468
Db 419 qrsdvgyyicqtlnvagisliitkaylevtdviadrppvtrgqpvqnvatvavdgfvlscva 478
Qy 469 TGNPSPRIKWFHDGHAVQA-GNRYSIQSSSLRVDLQLSDSGTYTCTASGERGETSWAA 527
Db 479 tqspvptilwrkdgvltvstqdsrikqlengvliqiryaklgdtgrytciastpsgeatwsa 538
Qy 528 TLTVKEPG-STSLHRAADPSTYPAPPGTPKVLNVSRTSISLRWAKSQEKGAVGPIIGYT 586
Db 539 yievqefgvvqpprptdpnlipsapskpevtvsrntvtlsw---qpnlnsgatptsyi 595
Qy 587 VEYFSPDLQTGWIVAHRVGDQTQVTSIGLTPGTSYVFLVRAENTQGISVPSGLSNVIKI 646
Db 596 ieafshasgsswgtvaenvktetsaiklqkpnaiylflvraanaygisdpqsdgvktq 655

Qy 647 EADFDAASANDLSAARTLLTGKSVELIDASAINASAVRLEWMLHVSADKYEGLRIHYK 706
Db 656 dv-lptsqgvdhkqvqrelgnavlhlhntplsssssievhwt--vdqqsqyigqykilyr 712
Qy 707 DASVPSAQYHS-----ITVMDASAESFVVGNLKKYTKYEFFLTPFFETIEGQPSNSKTA 760
Db 713 ---psganhgdesdwlfevrtpaknsvipdlrkvgvnyekarpffnefggadseikfa 768
Qy 761 LTYEDVPSAPPDNIQIGMY--NQTAGWVRWTPPSQHNGNLGYKIEVSAGNTMKVLAN 818
Db 769 ktleapsappqgvtskndngntailvswqpppedtqngmvqeykv-wclnetryhin 827
Qy 819 MTLNATTSVLLNNLTGAVISVRLNSFTKACDGPYSKPISEMDPTHVHPRAHPSGT 878
Db 828 ktvdgstfsvvipflvpglrysveaastgagsgvkspeqfiql-----ah---- 874
Qy 879 HDGRHEGQDLTYHNGN-IPPGDINPTTHKKTIDYLSGP-----WMLVLCIVLL 927
Db 875 -----gnvpsped-qvslaqqisdvkvqafiaagigaacwillnvfsiwl 918
Qy 928 VLIVISAISMVYFKRKHQ-----MTKELGHLSSVSDNEITALNI 966
Db 919 -----yrhrkkrngltstyagirkvpsftftptvtyqrggeavssggrrplni 967
Qy 967 NKSESL-WI-----DHHGRWRTADTDKDSGLSESLLSHVNSSQ--SNYNNS----- 1010
Db 968 sepaagpwladtwpntgnnhndcsiscotagnngsdnlttysrpadcianyannqldnkq 1027
Qy 1011 ----DGGTDYAEVDTRNLTTFFYNCRKSPD-----NPTPYATTMIIGTSSSETC 1054
Db 1028 tnmlpestvygdvldnskinemktfnsplkdgrfvnpsgqpttqqlgnslnnm 1087
Qy 1055 TKTTISISADK-----DSGTH 1069
Db 1088 nngsgdsgekhkwplgqgkqevapvqyniveqnlkndydrandvtvptipynqsydqntg 1147
Qy 1070 SPYSDAFAGQVPAVPPVKNYLYQYVPEP---INWSEFLPPPEHPPPSSTYGYAQGSPE 1125
Db 1148 gsynssdrgstsgsgghkkgartpkvpkgmgnwadllppphahpphs----- 1197
Qy 1126 SSRKSSKAGSGISTNQSIILNASIHSSSSGGFSAGWVSPQYAVACPPENVY----- 1176
Db 1198 -----nseeynisvdes-----ydgempcpvpparmylqqdeleee 1233
Qy 1177 -----SNPLSAVAGGTQNRQYITPTNQH--PPQLPAYFATTPGPGGAVPPNH 1220
Db 1234 edergtpvprgaasp-aavysyshqstatltspqaelqmlqdcpeetg-----h 1284
Qy 1221 LPFATQRAASEYQAGLNAARCAQSRACNSCDALATPSPMPPPPVPVPEGVQPVHPNS 1280
Db 1285 mqhqpdr-----qpvspppp-prp---ispptyg 1312
Qy 1281 HPMHPTSSN 1289
Db 1313 ylsqplvsd 1321

RESULT 6

Y13565

ID Y13565 standard; Protein; 1297 AA.

XX

AC Y13565;

XX

DT 30-JUL-1999 (first entry)

XX

DE C. elegans Robo polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commisureless; roundabout;
modulation; nerve cell function.

XX

OS Caenorhabditis elegans.

XX

PN WO925833-A1.

XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.

CC The invention relates to a method for modulating the amount of Comm
CC (commisurreless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 1297 AA;

Query Match 21.4%; Score 1588; DB 20; Length 1297;
Best Local Similarity 31.0%; Pred. No. 3.8e-83;
Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

Qy 55 SPRIIEHPTDLVKKNEPATLNCKVEGKPEPT-IEWFKDGEVPSTNEKK--SHRVQFKDG 111
Db 29 apvliehpidvvsrgspatln--gakpstakitwykgdpvitnkeqvshrlvdtg 86
Qy 112 ALFFYRTMQGK--KEQDGGGYWCVAKNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKG 169
Db 87 slfillkvnsgkngkdsdagaycvasnehevknsnegsklamlredfrvrprtvalgg 146
Qy 170 ETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNLLISNVEPIDEG 229
Db 147 emavlecspprgfepvsvwrkdd-----kelriqdmprylthsdgnllidpvdrsdsg 200
Qy 230 NYKICIAQLVGTRESSYAKLIVQVYFMEKPKDQVMLYGTATPHCSVGGDPPPKVLWK 289
Db 201 tygcvanmvgervsnparlsvfekpkfegqkdmtdvgaavldfcrtvgdppqgtwk 260
Qy 290 KEEGNIPVSRARILHDEKSLISNITPTDGTYYCEAHNNVQGISARASLIVHAPPNFTK 349
Db 261 rknepmvtrayiakdnrglriervqpsdegeyvcyarnpagtleasahlrvqappsfgt 320
Qy 350 RPSNKKVGLNGVVQLPCMASGNPPSVFVWTKGVSTLMFPN--SSHGRQYVAADGTLQIT 407
Db 321 kpadqsvpaggtatfectlvqgsppayfwakegqddllfpysadgrtkvsgptgtltie 380
Qy 408 DVRQDEGYVVCSAFVSVDSTVRVFLQVS----- 437
Db 381 evrqvdegayvcagmsagsslskaalkatfetkqrqvqkkskmgkqkqkvqsiikyli 440
Qy 438 -----SVDERPPPIIQIGPANQTLPGKSVATLPCRATGNPSPIKWFHDGHAQV-AGNRY 491
Db 441 savtgntpakppptiehgqntlmvgssailpcqasgkptgswlrdglpididsri 500
Qy 492 SIIQSSSLRVDLQLSDSGTYTCTASGERGETSWAATLTVEKPGSTS-LHRAADPSTYPA 550
Db 501 sqhstgslhiadlkpdtgvytciakndgestwsaltvedhtsnaqfvrmpdpnsfns 560
Qy 551 PPGTPKVLNVSRTSISLRWAKSQEKGAVGPIIGYTVVEYFSPDLQGTGWIYAHRVGDQV 610

Db 561 sptqplivnvtdevelhw--napstsgagpitgyliqyyspdigqtwfnipdyvastey 618
Qy 611 TISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIEADFDAASAN---DLSAARTLLT 666
Db 619 rikglkphsymfviraenekgigtvssalvttskpaagvalsdknmdmaiaekrlt 678
Qy 667 GKS-VELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYK-DASVPSAQYHSITVMDAS 724
Db 679 seqlikleevktinstavrlfwkkrkl--eelidgyyikwrgprrtndnqy--vvnvtsps 734
Qy 725 AESFVVGNLKKYTKYEFPLTPF---FETIEQGPSNSKTALTYEDVPSAPPDNIQIGMNYQ 781
Db 735 tenyvvsnlmpftnyeffviphysghsihgapsnsmdvltaeappsldpvrirmlnl 794
Qy 782 TAGWVRWTPPPSQHNGNLYGYKIEVSAGNTMKVLANMTLNATTSVLLNNLTGAVYSV 841
Db 795 ttrlskwpakdggingilkqfqi-vivgqapnnrnttneraasvltfhlvtgmytki 853
Qy 842 RLNSFTKAGDPYSKPSISLEMD-PTHVHPRAHPSGTHDGRHGGDLTYH--NNGNIPP 898
Db 854 rvaarsngvgvshgtsevimmqdtlekhl-----aqgenesflylinkshvp- 903
Qy 899 GDINPTTHKKTIDYLSGPNLMLVLCIVLLVLVISAISMVYFKRKHQMTKELGHLVSVSD 958
Db 904 -----vivivailiifvliiaycywrnsrnsdgdksrfsikind 942
Qy 959 NEITALNINSKESLW-----IDHHRGWRTADTKDSEKLLSHVNSQSNYNNSD-- 1011
Db 943 gsvhmasnn-----lwdvaqnpnqmpmyntagrtmnnrnngqalysitpnaqdfnncddy 998
Qy 1012 GGT-----DYAEV---DTRNLTFYNCRKSPDNPTPYATTMIIGTSSSETCTKTT 1058
Db 999 sgtmhrpgsehhyhaqltggpgnamstfyg-nqyhdpspyatttlv----- 1045
Qy 1059 SISADKDSGTHSPYSDAFAGQVPAVVPVKSNNLYQYVPEPINWSEFLPPPPEHPPSSYTG 1118
Db 1046 -----lsnqqa--wldnkmlrapamptn-----pvppe--pparyad 1079
Qy 1119 YAQGSPESSRSKSSKAGSGISTNQSLNASIHSSSSGGFSANGVSPQYAVACPPENVYSN 1178
Db 1080 htay--rrsrssrasdgrg-----tlngglhrtsgsgrs-----dsphtdvsy 1122
Qy 1179 PLSAVAGGTQNRQITPTNQHPQLPAY-FATTPGGAVPP-NHL-PFATQRHAASEYQA 1235
Db 1123 vqlhssdgtgskkertgerrtpnktlmdfippppsnpppgghvydtatrrq----- 1175
Qy 1236 GLNAARCAQSACNSDALATPSMQPPPPVPEGWYQVHPNSHPMHTSSNHQIYQC 1295
Db 1176 -lnrgstpredtyds-----vsdgafarvdna---rptsrnrnl--- 1211
Qy 1296 SSECSDHSRSSQSHKRLQLEEHGSSAKQROGGHRRRAP 1334
Db 1212 ggrplkgkrdddsqrslmdddggsseadgensegdvdp 1250

RESULT 7
Y08403
ID Y08403 standard; Protein; 1297 AA.
XX
AC Y08403;
XX
DT 24-JUL-1999 (first entry)
XX
DE C. elegans ROBO protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Caenorhabditis elegans.
XX
PN W09920764-A1.
XX

PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57252.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 59-63; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1297 AA;

Query Match 21.4%; Score 1588; DB 20; Length 1297;
Best Local Similarity 31.0%; Pred. No. 3.8e-83;
Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

Qy 55 SPRIIEHTDLVVKNEPATLNCKVEGKPEPT-IEWFKDGEVSTNEKK--SHRVQFKDG 111
Db 29 apvilehpidvvsrgspatln--gakpstakitwykdqgptvlnkeqvnshrivldtg 86

Qy 112 ALFFYRTMQGK--KEQDGGYWCVAKNRVQAVSRHASLQIAVLRRDPRVEPKDTRVAKG 169
Db 87 sifllkvnsgkngkdsdagayycvasnehgevknsyklamlredfrvrprtvtqalg 146

Qy 170 ETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNNLISNVEPIDE 229
Db 147 emavlecpgrgpepvsrwrkdd-----kelriqdmprytlhsdgnlidpvdrsdg 200

Qy 230 NYKCIANLVGTRESSYAKLIVQKPKPMKEPKDQVLMYLGQTATFHCVGGDPPPKVLWK 289
Db 201 tyqcvanmmvgervsnparslvfekpkfegepkdmtdvgaavldcrtvtdgppqitwk 260

Qy 290 KEEGNIPVSARILHDEKSLEISNITPTDEGTVCYCAHNNVGOISARASLIVHAPPNFTK 349
Db 261 rknepmvtraylakdnrglriervqpsdegeyvcyarnpagtleasahlrvqapsft 320

Qy 350 RPSNKKVGLNGVQLPCMASGNPPSVFVWTEGVSITLMPFN--SSHGRQYVAADGTLQIT 407
Db 321 kpadqsvpaggtatfectlvqgspayfwskeggqdlfpysvsadgrtkvsptgtltie 380

Qy 408 DVRQDEGYGYVCASFVSDSSTVRVFLQVS----- 437
Db 381 evrqvdegayvcagmsagsslskaalkatfetkgrvqkkskmgkqkqnvqsiyki 440

Qy 438 -----SVDERPPPIIGPANQTLPGKSVATLPCRATGNPSPRIKWFHDHAVQ-AGNRY 491
Db 441 savtgntpakppptiehgqngtlmvgssailpcqasgkptpgiswlrldgpliditsri 500

Qy 492 SIIQSSSLRVDDLQSDSGTYTCTASGERGETSWAATLVEKPGSTS-LHRAADPSTYPA 550
Db 501 sqhstgshliadlkpdtgvtyctiaknedgestwsasltvedhtsnaqfvrmpdpnsf 560

Qy 551 PPGTPKVLNVSRTSISLRWAKSQEKGAVGPIIGTYVEYFSPDLQIGWIAAHRVGDQV 610
Db 561 sptqpiivnvtdevelhw--napstsgagpitgyiiqyspdlqgtwfnipdyvastey 618

Qy 611 TISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIEADFDAASAN---DLAARTLLT 666
Db 619 rikglkpsysymfviraenekigtptsvssalvttskpaqvalsdknmdmaiaekrit 678

Qy 667 GKS-VELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYK-DASVPSAQYHSITVMDAS 724
Db 679 seqlikleevktinstavrifwkkrl--eelidgyyikwrgpprtndngy--vvnvtsp 734

Qy 725 AESFVVGNLKKYTKYEFLLTPF---FETIEGQPSNSKTALTYEDVPSAPPDNIQIGMNYQ 781
Db 735 tenyvvsnlmpftnyeffvipyhgsvhsihgapsnmdvltaeappsppedvrirmln 794

Qy 782 TAGVWRWTPPPSQHNGNLGYKIEVSAGNTMKVLANMTLNATTSVLLNNLTGAVYSV 841
Db 795 tlriskwkapkadgingilkfqi-vivgqapnnrnttneraasvtfhlvtgmytki 853

Qy 842 RLNSFTKAGDGPYSKISLPMF-PTHVHPRAHPSGTHDRHGQDLTYH--NNGNIPP 898
Db 854 rvaarsnggvvshgtsevimmqdtlekhla-----aqgenesflyglinkshvp- 903

Qy 899 GDINPTTHKTKTDYLSGFWLWLVLCIVLLVLSAISMVYFKRKHQMTKELGHSVSD 958
Db 904 -----vivivallifvliiaycywnrsnrsdgkdrsfikind 942

Qy 959 NEITALNINSKESLW-----IDHHRGWRTADTKDSEKSLSHVNSQSNYNNSD-- 1011
Db 943 gsvhmasnn---lwdvaqnpqnpmyntagrmtmnrrngqalysltpnagdfnnccdy 998

Qy 1012 GGT-----DYAEV---DTRNLTTFYNCRKSPDNPTPYATTMIIGTSSETCTKTT 1058
Db 999 sgtmhrpgsehhyhaqltggpgnamstfyg-nqyhdpspyatttlv----- 1045

Qy 1059 SISADKSGTHSPYSDAFAGQVPAVPVVKSNYLYQVPEPINWSEFLPPPPEHPPPSSTYG 1118
Db 1046 -----lsnqqa--wlndkmlrapamptn-----pvppe--pparyad 1079

Qy 1119 YAQGSPESSRSKSSAGSGISTNQSLNASIHSSSSGGFSANGVSPQYAVACPPENVYSN 1178
Db 1080 htay--rrsrssrasdgrg-----tlnglhlrtsgsgrs-----dsphtdvsy 1122

Qy 1179 PLSAVAGGTQNRQITPTNQHPPQLPAY-FATTGPGGAVPP-NHL-PFATQRHAASEYQA 1235
Db 1123 vqlhssdgtgsskertgerrtpnktlmdfippppsnpppgghvydtatrq----- 1175

Qy 1236 GLNAARCAQSRACNSCDALATPSMPQPPVPVPEGVQPVHPNSHPMHTSSNHQIYQC 1295
Db 1176 -lrgstpredtyds-----vsdgafarvdvna---rptsrnrnl--- 1211

Qy 1296 SSECSDHSRSSQSHKRLQLEEHGSSAKQSGHRRRAP 1334
Db 1212 ggrplkgkrddsqsslmdddgssadgensegdpv 1250

RESULT 8
Y08404
ID Y08404 standard; Protein; 1649 AA.
XX
AC Y08404;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
XX
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.

```

RESULT  9
W83927
ID  W83927 standard; Protein; 753 AA.
XX
AC  W83927;
XX
DT  01-MAR-1999 (first entry)
XX
DE  Human T85 protein.
XX
KW  T85; FHMB-6D4; FMHV-SD4; human; neurological disorder; therapy;
KW  diagnosis.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  Peptide 1..20
FT  /label= Sig_peptide
FT  Protein 21..753

```

```

FT      /label= Mat_protein
FT      Region      525..610
FT      /note= "has homology to a fibronectin type III
FT      domain"
FT      Region      638..727
FT      /note= "has homology to a fibronectin type III
FT      domain"
FT      Region      43..101
FT      /note= "has homology to a Ig superfamily domain"
FT      Region      145..203
FT      /note= "has homology to a Ig superfamily domain"
FT      Region      237..298
FT      /note= "has homology to a Ig superfamily domain"
FT      Region      329..394
FT      /note= "has homology to a Ig superfamily domain"
FT      Region      433..491
FT      /note= "has homology to a Ig superfamily domain"
FT      Peptide      247..249
FT      /note= "RGD motif"
FT      Domain      516..600
FT      /note= "cytokine receptor homology N-terminal
FT      domain"
IN      WO9848051-A2.
XX      :
PD      29-OCT-1998.
XX
PF      17-APR-1998;   98WO-US07714.
XX
PR      10-OCT-1997;   97US-0062017.
PR      18-APR-1997;   97US-0044746.
XX
PA      (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI      Holtzman D,  McCarthy SA;
XX
DR      WPI; 1999-024021/02.
DR      N-PSDB; V69278.
XX
PT      New isolated human FTHMA-070 and T85 proteins - used to develop
PT      products for the diagnosis and therapy of disorders involving
PT      cellular processes; e.g. neuronal development.
XX
PS      Claim 31; Fig 3; 127pp; English.
XX
CC      This is the amino acid sequence of a novel human protein designated
CC      T85, and also referred to as FMHB-6D4 and FMHB-SD4. T85 cDNA (see
CC      V69278) was identified in a human foetal brain cDNA library using a
CC      screen designed to identify genes encoding proteins having a
CC      functional signal sequence. T85 nucleic acids and polypeptides of
CC      the invention are useful as modulating agents in regulating a
CC      variety of cellular processes. They can be used for identifying
CC      compounds which bind to or modulate the activity of the polypeptides
CC      (claimed). They can also be used in screening assays, detection
CC      assays (e.g. chromosomal mapping, tissue typing, forensic biology),
CC      predictive medicine (e.g. diagnostic assays, prognostic assays,
CC      monitoring clinical trials, and pharmacogenomics), and methods of
CC      treatment (e.g. therapeutic and prophylactic) e.g. for neurological
CC      disorders.
XX
SQ      Sequence 753 AA;

```

Query Match 17.7%; Score 1317.5; DB 20; Length 753;
Best Local Similarity 39.0%; Pred. No. 6.5e-68;
Matches 284; Conservative 119; Mismatches 288; Indels 37; Gaps 15;

```
Qy      56 PRIIEHPTDLVVKKNENPATLNCKVVEGKPEPTIEWFKDGEVPVST--NEKKSHRVQFKDGAL 113  
       |||:||||:|:|:| | | | | | | :| :| | | | | :| :| :| |:| :|  
Db      29 privehpsdlivskgepatlnckaegrptptiewykggervetdkddprshrmllpsgsI 88  
  
Qy     114 FFYRTMQGKKEQ-DGGEYWCVAKNRNVQAVSRHASLIQIAVLRDDFRVEPKDTRVAKGETA 172
```

```

Db      89 fflrivhgksrpdgyvvcvarnylgeavshnaslevalldrrqndspvmvavgpa 148
Qy     173 LLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFAGASSVRIVDGGNLLISNEVIDEGNYK 232
       :||| |||: ||||: | || ||||           :| :| ||:|
Db     149 vmecqpgrghpeptiswkkdgspldd-----kderiti-rggklmityrksdagkyv 200
Qy     233 CIAQNVLGTRRESSYAKLIVQVKPYFMKEPKDQVMLYGQTATFHCSSVGDDPPPKVLWKKEE 292
       |: |:| ||| |:| |:| |:| |:| :   :| |:| ||| |:|::
Db     201 cvgtnmvgereasevaeltlvlerpsfvkrpsnlavtvdbsaefkceargdpvytvrwrkdd 260
Qy     293 GNIPVSRRARILHDEKSLEISNITPTDEGYTVEAHNNVGOISARASLIVHA---PPNFTK 349
       |:| :| |:| |:| :| |:| |:| |:| |:| |:| |:| |:| |:|
Db     261 gelpksryei-rddhtlkirkvtagdmgsytcaenmvgkaeasatlvtqvgssepvhfv 319
Qy     350 RPSNKVKGLNGVVQLPCMASGNPPSPVFVTKEGVSTILMF---PNSSHGRQYVAADGTLOI 406
       :| :: |:| |||: |:| ||| |:| |:| |:| |:| |:| |:| |:|
Db     320 kprdqvvalgrttvtfqeatgnppaifvrreogsnllfsypppgssrsfvsqtgdlti 379
Qy     407 TDVRQEDEGYVCSAFSVDSSTVRVELQVSSV-DERPPPIIQIGPANQTLPKGSVATLP 465
       |:|:: ||| |:| :| :| |:| |:| |:| |:| |:| |:| |:|
Db     380 tnvqrsdygyiqctlnvagsiitkaylevtdviadrpppvirggpvngtqvavdgtfvl 439
Qy     466 CRATGNPSPRIKFWDHGHAQVA-GNRYISIQQSSSLRVDDLQLSDSGTYTCTASGERGETS 524
       ||| |:| |:| ||| :| :| :| :| :| |:| |:| |:| |:|
Db     440 cvatgpsvpptilwrkdgvlstqdsrikqlengylgiryaklgdtgryctiaastpsgeat 499
Qy     525 WAATLTIVEKPG-STSLHRAADPSTYPAPPGTPKVNLVSRSTSISLRWAKSQEKPGAVGPPII 583
       |:| :| ::| | ||| :| :| |:| |:| ::| |:| :| :|
Db     500 wsayleyqefgypvqprrptdnpnlipsapskpevtwsrntvtls---qpnlinsgatpt 556
Qy     584 GYTVEYFSPDLQTGWIVAHRVGDITQVTISGLTPGTSYVFLVRAENTQGISVP SGLSNVI 643
       |:| |:| :| :| | | ||| |:| ||| |:| |:| |:| :| :|
Db     557 sylleafshagsgsswtvaenvktsaiklykpnaiylflvraanaygisdsqsisdcpv 615
Qy     644 KTIEADFDAASANDLSAARTLLTGKSVELIDASAINASAVRLEWMLHVSADERKYVEGLRI 703
       || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     617 ktqdv-lptsyygdvhkqwqrrelgnavlhlhntptvlsesslevhw-tvdqgsqiyigyki 773
Qy     704 HYKDASVSPAQYHS-----ITVMDASAESFVVGNLKKYTKYEFFLTTPPETIEGQPSNS 757
       |:| :| | | :| :| |:| |:| ||| |:| :| :|
Db     674 lyr-----psganhgesdwlfvevrtpaknsvvlpdlrkgvnyeikarpffnefggadsei 729
Qy     758 KTAITYED 765
       ||| |:|
Db     730 kfaktlee 737

```

```

RESULT 10
W42087
ID W42087 standard; Protein; 1571 AA.
XX
AC W42087;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human Down syndrome-cell adhesion molecule DS-CAM2.
XX
KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis; assay; human.
XX
OS Homo sapiens.
XX
PN W09817795-A1.
XX
PD 30-APR-1998.
XX
PF 23-OCT-1997; 97WO-US19547.
XX
PR 25-OCT-1996; 96US-0029322.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.

```

XX
PI Korenberg JR;
XX
DR WPI; 1998-271791/24.
DR N-PSDB; V31988.
XX
PT New isolated Down's Syndrome-cell adhesion molecule - used to
PT develop products for detection, diagnosis and therapy of
PT developmental and neurological abnormalities
XX
PS Claim 2; Page 90-95; 109pp; English.
XX
CC This polypeptide comprises Down syndrome-cell adhesion molecule
CC DS-CAM2, an extracellular soluble protein belonging to a novel
CC subclass of the Ig superfamily with highest homology to neural cell
CC adhesion molecules. Its amino acid sequence was deduced from cDNA
CC clones (see V31982) isolated from a trisomy 21 foetal brain library.
CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and
CC lacks the entire transmembrane domain of DS-CAM1. The invention
CC provides human and murine DS-CAM nucleic acid sequences (see also
CC V31981, V31985-87), expression vectors and host cells, transgenic
CC animals, antibodies, antisense oligonucleotides, and primers
CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
CC with developmental and neurological processes. They can be used in
CC e.g. neural prosthetic devices used in entubulation methods of
CC repairing (regenerating) damaged or severed peripheral nerves, and
CC also in bioassays to identify agonists and antagonists. The products
CC can also be used in detection, diagnosis and therapy of developmental
CC and neurological abnormalities such as Down syndrome, mental
CC retardation, holoprosencephaly, agenesis of the corpus callosum,
CC or schizencephaly.
XX
SQ Sequence 1571 AA;

Query Match 9.04; Score 665.5; DB 19; Length 1571;
Best Local Similarity 23.84; Pred. No. 5.8e-30;
Matches 322; Conservative 184; Mismatches 492; Indels 353; Gaps 65;

Qy 55 SPRIIEHPTDLVKKNEPATLNCKVEGKPEPTIEWFKDGEPEVSTNEKSHRVQ---FKDG 111
Db 406 tpkliisafsekvvpaepvslmncvkgtpitwltiddpdl---kqgshrisqmitseg 463
Qy 112 ALFFYRTMQGKKEQGGGYWCVAKNRVQAVSRHASLIQIAVLRRDFFRVEP-KDTRVAKGE 170
Db 464 nvlslinissqvrddggyrctannasag-vvlyqarinv---rgpasirpmknitaiagr 519
Qy 171 TALLECGPPKGIPEPTLIWIKDG--VPLDDLKAMSGASSRVRIVDGGNLLISNVE-PID 227
Db 520 dtyihc-rvigypysikwyknsnllpfn-----hrqvafengntlklsdvqkev 569
Qy 228 EGNKYC--IAQNLVGTRESSYAKLVQVKPYF--MKEPKDQVMLYGQTATFHC-SVGGDP 282
Db 570 egeytcnvlvqpqlstsgsvh--vtvkvppfiqpfefprfsi---gqrvfipcvvysgd 624
Qy 283 PPKVLWKKKEGNIPVSRARILHD---EKSLEISNITPTDECTVCEAHNVGQISARSL 339
Db 625 pititwqkdgrpipslgvtdnidftsslrslmhnqnytciaarneaaavehqsq 684
Qy 340 IVHAPPNFTKRPSNKKVGLNG-VVQLPCMASGNPPSVFVT-KEGVSTLMF-PNSSHGRQ 396
Db 685 ivrvppkfivvqprdq-dglykavilncsaegypvptivwfksgagvqpqfpialngri 743
Qy 397 IVAADGTLQITDVRQDEGYVCSAFVSVDSTVR-VFLQVSSVDERPPPIIIGPANQT 455
Db 744 qvlsngsllkhvveedsgyyllckvsnvdgadvksmyltv----kipamitsypnttl 798
Qy 456 LPKGSVATLPCRATGNPSPRIKWFHDGHAVQAG-NRYSIIQG-----SSLRVDDLQLS 507
Db 799 atqgqkemsctahgekiipivrewkedriinpemarylvstkeveevistlqilptvre 858
Qy 508 DSGTYTCTASGERGETSWAATLTVEKPGSTSLHRAADPSTYPAPPGTPKVLNVSRTSISL 567

Db 859 dsgrffschainsygedrgiiqltvqep-----pdppei-eikdvkartitl 903
Qy 568 RWAQSQEKPGAVGPIIGYTVBYFSPDLQGWIVAHRVGD-----TQVTISGLTPGTSYV 622
Db 904 rwtmgfd---gnsptitgydie--cknksdsw-dsaqrtdkvpqnsatidihpsstys 957
Qy 623 FLVRAENTQGISVPSGLSNVKTIEADFAASANDLSAARTLLTGKSELIDASAINASA 682
Db 958 imyaknrigksep---snel-titad-eaap-----dgpqqe-vhlepiissqs 1000
Qy 683 VRLEWMLHVSADKEY-----VEGLRIHYKSDASVPSAQYHSITVMDASAES--FVVGNLKK 735
Db 1001 lrvtw---kapkhlqngilrirygygyreystggnfqniisvtdsgsdsevytldnlnk 1056
Qy 736 YTKYEFFLTPTFFETIEGQSPNSKALTIEDVPSAPPDNIQIGMYNQTAGVWRTPPSQ 795
Db 1057 ftqyglvqvacnragtgpssqeittledvpsyppevqaiatspesisiswstlskea 1116
Qy 796 HNGNLGYKIEVSAGNTMKVLNMTLNATTT--SVLLNNLTGAVISVRLNSFTKAGDGP 853
Db 1117 lngilqgfrv-lywanlmdgelgeiknltttqpsleldglekytnysiqvlfatragdv 1175
Qy 854 YSKPISLFMDETHHVHPPRAHPSGTHDRHEGQDLTYHNNGNIPPGDINPTTHKTTDYL 913
Db 1176 rseqi---ftrtkedvpgp---pag----- 1194
Qy 914 SGPWMLVLCIVLLVLVISAAMVYFK-----RKHMKTGELHLSVVDNE-- 960
Db 1195 -----vkaaaasasmvfvswlpplkngilrkytvcfshpyptvisfeas 1240
Qy 961 -----ITALNINSKESLWIDHHRGWRTADTDKDSLSESKLLSHVNSQSNNNSDGG 1013
Db 1241 pdsfsyripnlrnrqysvww-----vavtsagrg-----nssei 1276
Qy 1014 T--DYAEVDTRNLT-----TFYNCRKSPDNPTPYATTM--IIGTSSSETCTKT 1057
Db 1277 tveplakapariltfsgtvtppwmkdivlpc-kavgdpspavkwmkdsngtspvltdigr 1335
Qy 1058 TSISAD-----KDSGTHSPYSDAFAGQVPAVPVVKSNYLQYVPEPINWSE---- 1102
Db 1336 rsifsgsfliirtvkaedsgyys-----ciann-----nwgdsdeill 1372
Qy 1103 --FLPPPEHP-----PPSSTYGYAQSGSPSSRKSAGSGISTNQSLNASIHSS 1152
Db 1373 nlqvqvpdpqorltvsktssitlswlpqd-----nggssirgyilqysedns 1421
Qy 1153 SSGGFSAWGVSPQYAVACPPENVSNPLSAVAGGTQNRQITPTNQHP----- 1201
Db 1422 eq-----wgsfp---ispersyr--lenlkcgtwykftltaqngvpggriseieakt 1470
Qy 1202 -----QLPAYFATT-----GPGG-----AVPNHLPFAT-----QR---- 1227
Db 1471 lgkepqfskeqelfasinttrvrlnlgwdggcpitsftleyrpfgtvtwtaqrtsls 1530
Qy 1228 -----HAASEYQAGL---NAARCAQSRA 1247
Db 1531 ksyilyldheatwyelqmrvcnsagcaekga 1561

RESULT 11

W42086

ID W42086 standard; Protein; 1910 AA.

XX

AC W42086;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM1.

XX

KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis; assay; human.

XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= Sig_peptide
FT Protein 24..1910
FT /label= Mat_protein
FT Domain 24..887
FT /label= IG
FT /note= "immunoglobulin type-C2 domain"
FT Domain 888..1594
FT /label= FbN
FT /note= "fibronectin type III domain"
FT Domain 1595..1616
FT /label= Transmembrane
FT Domain 1617..1910
FT /label= Cytoplasmic
FT Region 24..126
FT /label= Ig1
FT Region 127..225
FT /label= Ig2
FT Region 226..316
FT /label= Ig3
FT Region 317..409
FT /label= Ig4
FT Region 410..506
FT /label= Ig5
FT Region 507..603
FT /label= Ig6
FT Region 604..697
FT /label= Ig7
FT Region 698..792
FT /label= Ig8
FT Region 793..887
FT /label= Ig9
FT Disulfide-bond 46..102
FT Disulfide-bond 145..197
FT Disulfide-bond 246..293
FT Disulfide-bond 335..385
FT Disulfide-bond 428..484
FT Disulfide-bond 525..575
FT Disulfide-bond 617..669
FT Disulfide-bond 711..766
FT Disulfide-bond 809..865
FT Disulfide-bond 1307..1359
FT Modified-site 78..80
FT /note= "Asn is N-glycosylated".
FT Modified-site 106..108
FT /note= "Asn is N-glycosylated"
FT Modified-site 470..472
FT /note= "Asn is N-glycosylated"
FT Modified-site 487..489
FT /note= "Asn is N-glycosylated"
FT Modified-site 658..660
FT /note= "Asn is N-glycosylated"
FT Modified-site 666..668
FT /note= "Asn is N-glycosylated"
FT Modified-site 710..712
FT /note= "Asn is N-glycosylated"
FT Modified-site 748..750
FT /note= "Asn is N-glycosylated"
FT Modified-site 795..797
FT /note= "Asn is N-glycosylated"
FT Modified-site 924..926
FT /note= "Asn is N-glycosylated"
FT Modified-site 1142..1144
FT /note= "Asn is N-glycosylated"
FT Modified-site 1160..1162
FT /note= "Asn is N-glycosylated"
FT Modified-site 1250..1252
FT /note= "Asn is N-glycosylated"
FT Modified-site 1271..1273

FT /note= "Asn is N-glycosylated"
FT Modified-site 1324..1326
FT /note= "Asn is N-glycosylated"
FT Modified-site 1341..1343
FT /note= "Asn is N-glycosylated"
FT Modified-site 1488..1490
FT /note= "Asn is N-glycosylated"

XX
PN WO9817795-A1.
XX
PD 30-APR-1998.
XX
PF 23-OCT-1997; 97WO-US19547.
XX
PR 25-OCT-1996; 96US-0029322.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Korenberg JR;
XX
DR WPI; 1998-271791/24.
DR N-PSDB; V31981.
XX
PT New isolated Down's Syndrome-cell adhesion molecule - used to
PT develop products for detection, diagnosis and therapy of
PT developmental and neurological abnormalities
XX
PS Claim 2; Page 73-78; 109pp; English.
XX
CC This polypeptide comprises Down syndrome-cell adhesion molecule
CC DS-CAM1, a cell surface glycoprotein belonging to a novel subclass
CC of the Ig superfamily with highest homology to neural cell adhesion
CC molecules. Its amino acid sequence was deduced from cDNA clones
CC (see V31981) isolated from a trisomy 21 foetal brain library. A
CC splice variant, DS-CAM2 (see W42087), which is non-membrane bound
CC was also identified. The invention also provides human and murine
CC DS-CAM nucleic acid sequences (see also V31985-88), expression
CC vectors and host cells, transgenic animals, antibodies, antisense
CC oligonucleotides, and primers derived from DS-CAM nucleic acid.
CC DS-CAM polypeptides are associated with developmental and
CC neurological processes. They can be used in e.g. neural prosthetic
CC devices used in entubulation methods of repairing (regenerating)
CC damaged or severed peripheral nerves, and also in bioassays to
CC identify agonists and antagonists. The products can also be
CC used in detection, diagnosis and therapy of developmental and
CC neurological abnormalities such as Down syndrome, mental
CC retardation, holoprosencephaly, agenesis of the corpus callosum,
CC or schizencephaly.
XX
SQ Sequence 1910 AA;

Query Match 8.9%; Score 661; DB 19; Length 1910;
Best Local Similarity 24.2%; Pred. No. 1.4e-29;
Matches 305; Conservative 175; Mismatches 473; Indels 308; Gaps 60;

Qy 55 SPRIIEHPTDLVVKKNPATLNCCKVEGKPEPTIEFWKDGPEVSTNEKSKSHRVQ---FKDG 111
Db 406 tkiisafsekvsvpaepvslmncvkgtpititwtdlddpil--kqgshrisqmitseg 463
Qy 112 ALFFYRTMQGRKEQDGGGYWCVAKNRVQAVSRHASLQIAVLRRDDFRVPE-KDTRVAKGE 170
Db 464 nvvsylnsssgvrdggvyrctannsag-vvlyqarinv---rgpasirpmknitalagr 519
Qy 171 TALLECGPPKGIPEPTLIWIKDG--VPLDDLKAMSFGASSRVRVDGNNLLISNVE-PID 227
Db 520 dtyihc-rvgyppysikwyknsnllpfn-----hrqvafenngtiklsdvqkev 569
Qy 228 EGNKYC--IAQNLVGTRESSYAKLIVQVKPYF--MKEPKDQVLMYQGTATFHC-SVGGDP 282
Db 570 egeytcnvlvqqlstsqsvh--vtvkvpfpfiqfefprfsi---gqrvfipcvvsvsgl 624
Qy 283 PPKVLWKKECNIPVSRARILHD---EKSLEISNITPTDEGTIVCEAHNNVQGISARASL 339

Db 625 pititwqkdgrpipslgvtidnidftsslrlnslmhnngnytciaarneaavehqsq 684
Qy 340 IVHAPPNFTKRPSNKKVGLNG-VVQLPCMASGNPPSPVEWT-KEGVSTLMF-PNSSHGRQ 396
Db 685 ivrvppkfvvqprdq-dgiygvkavilncsaegypvptivwfksgagvpgfqialnqri 743
Qy 397 YVAADGLQITDVRQDEGYVCSAFSVSDSTVR-VFLQVSSDERPPPIIQIGPANQT 455
Db 744 qvlsngsllikhvveedsgyyllckvsndvgadvksmyltv-----kipamitsypnttl 798
Qy 456 LPKGSVATLPCRATGNPSRIKWFHDGHAVQAG-NRYSIIQG-----SSLRVDLQLS 507
Db 799 atggqkckemstahgekipivrvkedriinpemarylvtstkevgvistlqlilptvre 858
Qy 508 DSGTYTCTASGERGETSWAATLTVEKPGSTSLHRAADPSTYPAPPPTKVLNVSRTSISL 567
Db 859 dsqffschainsygedrgiqltvgpe-----pdppel-eikdvkartitl 903
Qy 568 RWAKSQEKGAVGPIIGTYVEYFSPDLTGWIVAHRVGD-----TQVTISGLTPTSTV 622
Db 904 rwtngfd---gnsplitgydie--cknksdw-dsaqrtdkvsqplnsatidihpsstys 957
Qy 623 FLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSAARTLLTGKSVELIDASAINASA 682
Db 958 irmayknriksep---snel-titad-eaap-----dgppge-vhleipsisqs 1000
Qy 683 VRLEWMLHVSADKY-----VEGLRIHYKDVSAQYHSITVMDASAES--FVVGNLKK 735
Db 1001 irvtw----kapkhlqnglirgyqlygreystggnfqnliisvtdsgdsevytldnlnk 1056
Qy 736 YTKYEFFLTPFFETIEGQPSNKTALTYEDVPSAPPDNIQIGMYNQTAGVWRVTPPPSQH 795
Db 1057 ftqyglvqacnragtgpssqeitttleddvpsypnvgaiatspesisiswstlskea 1116
Qy 796 HNGNLYGYKIEVSAGNTMKVLNMTLNATTT--SVLLNLTGAVYSVRLNSFTKAGDGP 853
Db 1117 lngilgqfrv-iywanlmdgelgeiknitttqpsleldgelyktnysiqvlaftagdg 1175
Qy 854 YSKPISLFMDPTHVHPPRAHPSGTHDRHEGQDLTYHNNGNIPPGDINPTTHKKTIDYL 913
Db 1176 rseqi--ftrtkedvpgp---pag----- 1194
Qy 914 SGPWMLVLCIVLLVLVISAAMVYFK-----RKHMKTGELHSLVVSVDNE-- 960
Db 1195 -----vkaaaasasmvfswlplknglirkytvcfshpyptvisefeas 1240
Qy 961 -----ITALNINSKESLWIDHHRGWRTADTDKDSGLSEKLLSHVNSQSNYNNSDGG 1013
Db 1241 pdsfsyripnlrnrqysvww-----vavtsagrg-----nssei 1276
Qy 1014 T--DYAEVDTRNLT-----TFYNCRKSPDNPTPYATTM--IIGTSSSETCTKT 1057
Db 1277 tveplakapariltfsgtvtwpmkddivlpc-kavgdpsavkwmdsngtspstldgr 1335
Qy 1058 TSISAD-----KDSGTHSPYSDAFAGQVPAVPVVKSNYLYQVPEPINWSE---- 1102
Db 1336 rsifsgnsfiirtvkaedsgyys-----clann-----nwgsdeil 1372
Qy 1103 --FLPPPEHP-----PPSSTYGYAQGSPESSRKSAGSGISTNQSIILNAIHSS 1152
Db 1373 nlqvypdpqprltvsktssitlswlpgd-----nggssirgyilqysedns 1421
Qy 1153 SSGGFSAGWSPQYAVACPPENVSNPLSAVAGGTQNRQITPTNQHPPQLPAYFATTP 1212
Db 1422 eq-----wgsfp-----ispsersyr--lenlkcgtwykftltaqn-----gvgp 1459
Qy 1213 G 1213
Db 1460 g 1460

RESULT 12

R13144
ID R13144 standard; Protein; 1728 AA.
XX
AC R13144;
XX
DT 04-OCT-1991 (first entry)
XX
DE Deleted in Colorectal Carcinomas.
XX
KW DCC gene; cancer; diagnosis; antibodies; tumorigenesis; neoplasm.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 202..1648
FT /label= DCC
FT Peptide 202..227
FT /label= sig_peptide
FT Protein 228..1648
FT /label= mat_protein
XX
PN W09109964-A.
XX
PD 11-JUL-1991.
XX
PF 19-DEC-1990; 90WO-US07314.
XX
PR 04-JAN-1990; 90US-0460981.
XX
PA (UIJO) JOHNS HOPKINS UNIV.
XX
PI Vogelstein B;
XX
DR WPI; 1991-222913/30.
DR N-PSDB; Q12752.
XX
PT Human DCC gene, deleted in colorectal carcinoma - and diagnosis
PT or prognosis of neoplasms by detecting loss of gene function or
PT expression prods., or mutation(s)
XX
PS Claim 44; Page 31; 5ipp; English.
XX
CC Cells transformed with the wild-type DCC gene can be used as model
CC systems to study cancer remission and drug therapy. DCC polypeptide
CC expression prods. may be used to reverse the neoplastic state.
CC X1615 represent an amino acid illegible in the specification, all
CC other Xs are encoded by stop codons.
XX
CC See also Q12752:55.
XX
SQ Sequence 1728 AA;

Query Match 8.6%; Score 639; DB 12; Length 1728;
Best Local Similarity 23.1%; Pred. No. 2.2e-28;
Matches 361; Conservative 170; Mismatches 565; Indels 464; Gaps 70;

Qy 57 RRIEHTDLVVKKNPATLNCQVEG-KPEPTIEWFKDG--EPVSTNEKSHRVQFKDGL 113
Db 242 rflepdsdavnrmrgnvlldcsaesdrvgpvikwkdghlalgmderkq---qlsngsl 298
Qy 114 FFYRTMQGKKEQ-DGGEYWCVAK-NRVQAVSRHASLQIA-VLRDDFRVEPKDTRVAKGE 170
Db 299 liqnihlrhpkdeglyqceasldgsgsiistrtakvavagplr--flsqtesvtafmgd 356
Qy 171 TALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVIVDGGNLLISNVEPIDEGN 230
Db 357 tvllkc-ewigepmptihwqknqddltip-----gdsrvvvlpsgalqisrlpgdigi 410
Qy 231 YKCIQNLVGTRESSYAKLIVQVKP-----YFMKEPKDQVMLYGTATFHCSVGGDPPP 284
Db 411 yrcearnpassrtgneaevrilsdpghrlqlyflrpsnvvaiekgdavleccvsgyppp 470
Qy 285 KVLWKEEGNIPV-SRARILHDEKSLISNITPTDEGTYVCEAHNVGQISARASLIVHA 343

Db 471 sftwlrgeeviqrlskkysllggsnllisnvtddsgmytcvtyknenisasaeltvlv 530
Qy 344 PPNFTKRPSNKKVGLNGVVLPCMASGNPPSVFVTKEGVSTLMFPNSSHGQRQYVADGT 403
Db 531 ppwflnhpsnlyayesmdiefectvsgkpvptvnmkng--dvvipdyf--qivggssn 585
Qy 404 LQITDVRQEDEGYVCSAFSVSDSSTVRVFLQVSSVDERPPPIIIGPANQTLPGKSVAT 463
Db 586 lrilgvksdegfyqcvaeneagnaqtasqliv-----pkpaipsssvlpsaprdvvpv 639
Qy 464 L-----PCRTGNPSPRIKWFHDGHAVQAGNRSIIQ----GS-SLRVDDLQL 506
Db 640 lyssrfvrlswrpaekgn----igtftvfrsregdnreranttpqsglqvtgnlkp 695
Qy 507 SDSGTYTCTASGERGETSWAATLVEKPGSTS--LHRAADPS-TYPAPPPTKVLNVSRT 563
Db 696 eamtyfrvaynewg-----pgessqpkvatqpelqvgpgpvenlqavstps 743
Qy 564 SISLRWAKSQEKPG-AVGPIIGTYVEYFSPDLQTC----- 597
Db 744 sillitw---eppayangpvqgyrl--fctevstgkeqnievdqlyskleglkkfteysl 797
Qy 598 WIVAAHRVGD-----TQVTIS----- 613
Db 798 rflaynrygpgvstdditvtlsdvsappqnslevvnsrsikvswlpppsgtqngfit 857
Qy 614 -----GLTPGTSYVFLVRAENTQGISVPSGLSNVIRT 645
Db 858 gykirhrktrrgemetlepnlnlylftgleksgysfqvsamtvgtpgsnwyta-et 916
Qy 646 IEADFDAASANDLSAA-----RTLLTGKSV-----ELID 674
Db 917 pendldesqvpdpqsslhvrpqtncilmswtpplnpniwvrygylgyvgyspaetvrvd 976
Qy 675 ASAINASAVRLEWMLHV-----SADEKTV----- 698
Db 977 skqrysierlessshyvislkafnnagegvplyesatrsitdtpdvdypllddft 1036
Qy 699 -----EGLRIHYKDAASP-----SAQY 715
Db 1037 wvpdlstplmpvgqvalthdavrsvwadnsvpnkqktsevrlytvrtwsfsasaky 1096
Qy 716 HSITVMDASAESFVVGNLKKTKEFFLTPFFETIEGQPSNKTALTYEDVP-SAPPDNI 774
Db 1097 ks---edttlsytatgktpntmyefsvmtknrrsstwsmsahattyeaaptsapkdf 1153
Qy 775 QIGMYNQ-TAGWVRWTPPPSQHHGNLYGYKIEVSAGNTMK---VLANMTLNATTSTVL 829
Db 1154 vitregkpravivswpp--leangkitayilfytlkdnipiddwmetisgdrthqim 1211
Qy 830 LNNLTGAVYSVRLNSFTKAGDGPYKPSISLFMDPTHV-HPPR-AHPSGTH-DGRHEQ 886
Db 1212 dlnldt--myfriaqarnskgvpplsdpl-lfr--tlkvehpdkmandqgrhgdgywpv 1266
Qy 887 D--LTYHNNGNIPP-GDINPTTHKTTDYLSPGLMVLV---CIVLLVLVISAAI---- 935
Db 1267 dtnldtrstlneppigqmhphgsvtpqknsnllvllvvtvgvitvlvviavicttr 1325
Qy 936 SMVYFKRK---HQMTKELGHLVSVSDNEITALNINSKESLWIDHHRGRTADTKDSG-- 990
Db 1326 ssaqqrkrkthasagkrkg-----sqkdl-----rppdlwi-hheemenkiekpsgt 1373
Qy 991 -----LSESKLLSHVNSSQSNYNNSDGGTDYAEVDTRNLTFYCNCRKSPDNPTPIYAT 1043
Db 1374 pagrdsplqscqdltpvshsqsetqlgskstshsgqtee----- 1413
Qy 1044 MIIGTSSETCTKTTSISADK-----DSGTHSPYSDAFAGQVPAVPVVKSNLYQ 1093
Db 1414 -----agssmtlerslaarraprkrklmpmdagennp-----avsaipvptlesaqy 1462
Qy 1094 PVEPINWSEFLRPPP-EHPPSPSTY-----GYAQSPSSSRKSSKSAGSGIST 1140

Db 1463 p-----glpsptcgyphqftlrpvpfptlsvdrgfgag-----rsqsvsegptt 1508
Qy 1141 NQSIL---NASIHSSSSGGFSAWGVSPQYAVAC----PPENVYSNPLSAVAGGTQNYQI 1193
Db 1509 qppmlppsqspehssee-----apertiptacvrpthplrsfanplp----- 1552
Qy 1194 TPTNQHPPLPAYFATTGPGGAVPPNHLPFATQRHAASEYQAGLNAARCAQSRACNSCDA 1253
Db 1553 ppsalepkvpytpllsqpgptlpkthvk-----taslglagkars----- 1593
Qy 1254 LATPSMQPPPPVPVPEGWYQPVHPNSHPMHTSSNHQIYQCSSECDHSRSSQSHKRQL 1313
Db 1594 -----pllpvsypta--pevseesh--kptxdsanvye-qddlseqmasleglmkql 1640

RESULT 13

R68553

ID R68553 standard; Protein; 1447 AA.

XX

AC R68553;

XX

DT 05-JUL-1995 (first entry)

XX

DE Deleted in colorectal carcinoma (DCC).

XX

KW Tumour suppressor; deleted in colorectal carcinoma; antibody;
cancer diagnosis.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1..1063

FT /note= "DCC epitope"

FT Misc-difference 3369..4341

FT /note= "DCC epitope"

FT Misc-difference 26..1126

FT /note= "DCC epitope on extracellular domain"

FT Misc-difference 1123..1447

FT /note= "DCC epitope on intracellular domain"

XX

PN W09428161-A.

XX

PD 08-DEC-1994.

XX

PF 18-MAY-1994; 94WO-US05277.

XX

PR 26-MAY-1993; 93US-0068950.

XX

PA (UJJO) UNIV JOHNS HOPKINS.

XX

PI Bruskin A, Jarosz DE, Johnson K, Kinzler KW, Vogelstein B;
Zabrecky JR;

XX

DR WPI; 1995-022830/03.

DR P-PSDB; Q80196.

XX

PT Antibodies specific for tumour suppressor gene product, DCC -
useful for detecting expression of DCC gene, for cancer diagnosis.

XX

PS Claim 4; Page 24-28; 39pp; English.

XX

CC The protein represents the DCC tumor suppressor, and epitopes are
CC identified which are used in the generation of polyclonal or
CC or monoclonal antibodies against DCC. The antibodies can detect
CC DCC protein in biological samples (including tumour tissue,
CC peripheral blood mononuclear cells or a tumour biopsy lysate)
CC despite low levels of DCC expression, and are therefore useful in
CC cancer, especially colorectal carcinoma, diagnosis.

XX

SQ Sequence 1447 AA;

Query Match

8.6%; Score 637.5; DB 16; Length 1447;

Query Match . 8.6%; Score 637.5; DB 20; Length 1447;
Best Local Similarity 22.9%; Pred. No. 2.1e-28;
Matches 271; Conservative 150; Mismatches 444; Indels 317; Gaps 43;

```

Qy      35 WLLLVLVASNGLPAVRGQYQSPRIE-----HPTDLVVKKNPATLNCKVEGKPEPTI 87
      ||| : | : :| :|| : :| :|||
Db      9 wpll-lcspclllqipeeyeghhmveppviteqsprllvfvptdislkceasgkpevqf 67

Qy     88 EWFKDGPEVSTNEKSHRVQFKDGLFFPYRTMGKK---EQDGGEYNCVAKNRVQGAVS 143
      :|| :| : :| :| : :| :|| :|| :| :|
Db     68 rwrtdgwhfkpkeelgvtyvqspshsgsf--titgnnsnfagrfggyircfasnklgtams 125

Qy    144 RHASLQIAVLRDDFRVEPKDT---RVAKGETALLECGPPKGIPEPTLIWIKDGVPLDDL 199
      :| : :| :|| :| :| :| :| :| :|
Db    126 h-----eirlmdegapkwpketvkpveeegesvlpncpppsaeplrlywmns----- 174

Qy    200 KAMSGGASSRVIRVDGGNLLISNV-----EPI----- 226

```

Db 175 kilhikqdervtmgqngnlyfanvltsdnhsdyichahfpgttrtiqkepidlrkatns 234
Qy 227 ----- 226
Db 235 midrkprllfptnssshlvalqggplvleciaegfptptikwlrpsgmpadrvtynhn 294
Qy 227 -----DEGNYKCIAQNLVGTRESSYAKLIVQVKPYFMKEPKDQVMLYGQTATFH 275
Db 295 ktlqlkveeddgeyrclaenslgssarhay-yvtveaapywhkpgshlygpgetarld 353
Qy 276 CSVGDDPPPKVLWKKKEGNIPVSRARILHDEK-----SLEISNITPTDEGTYVCEAHNN 329
Db 354 cqvggprpgevtwr--ingipve--elakdqkyriqrgalilsnvqpsdtmvtgcearnr 409
Qy 330 VGQISARASL-IVHAPPNFTKRPSNKKVGLNG-VVOLPCMASGNPPSVFWTKEGVSTLM 387
Db 410 hglllanayiyvvqlpakiltadngtymavqgstayllckafgapvpsqwldeggtvl 469
388 FPNSSSHGRQYVAADGTLQITDVRQEDGEGYVCSAFSVVDSSTVRVFLQVSSVDERPPPII 447
Db 470 ----qderffpyangtlgirdlqandtgryfclaandqnnvtlmanlkvkdatg-----i 520
Qy 448 QIGPANQTLPGKSVATLPCRATGNPS--PRIKWFHDGHAVQA---GNRSIIQSSSLRVD 502
Db 521 tqgprstiekkgsrvtftcqasfdpslqpsitwrgdgrldgelgdsdky-fiedgrlvih 579
Qy 503 DLQLSDSGTYCTTASGERGET-SWAATLTVEKPGSTSLHRAADPSTYPAPPGTPKVLN-V 560
Db 580 sl dysdggnyscvasteldvvesraqlllvvgspg-----pvprrlvlsdlhl 626
Qy 561 SRTSISLRWAKSQEKPAGVPIIGYTVVEYFSPDL-QTGWIVAAHRVGDQTQVITISGLTPGT 619
Db 627 tqsqvrsvswpaedhn---apiekydiefedkemaekwyslgkvpngtsttlklspyv 683
Qy 620 SYVFLVRAENTQGISVPSGLSNVIRKIEADFDAASANDL-----SAARTLLTGKSVEL 672
Db 684 hytfrvtainkygpgespvssetvvtpea---apeknvdkvgegnettmmitwkplrw 740
Qy 673 IDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDASVPSAQYHSITVMDASAESFVVGN 732
Db 741 mdwnapqvq-yrvqwr-----pggtgpgwqeqiv-----sdpflvvsn 777
Qy 733 LKKYTKYEFLTPFFETIEGQPSNSKTALTY----EDVPSAPPDNIQIMYNTAGWVRW 788
Db 778 tstfvppeikv---qavnsqgkqepqvgtigysgedypqailegieilnssavlvkw 833
789 TPPPSQHNGNLYGYKIEVSAGNTMKVLA-----NMTLNATTISVLLNLTGAVYSV 841
Db 834 rpvdlagvkghlrgynvtywregsqrkhskrhikhdhvvpanttsvllsgrlpyssyhl 893
Qy 842 RLNSFTKAGDGPYSKPISLFMDPTH-HVHPPRAH-----PSGTHDGRHEGQ 886
Db 894 evqafngrgsgpase--ftfstpegvpghpealhlecqsntslllrwqplshngvltgy 951
Qy 887 DLYH 891
Db 952 vlsyh 956

Search completed: January 22, 2001, 12:17:03

Job time: 1580 sec

```

Qy 55 SPRIIEHPTDLVVKKNPATLCKVGEKPEPT--IEWFKDGEVPSTNEKK--SHRVQFKDG 119
   :| ||||| |:| : ||||| || | |:|:| | |: : | |:| |
Db 30 APVIEHPIDVVSRRGSPATLNC--GAKPSTAKITWYKDGQPVITNKEQVNSHRIVLDTG 87

Qy 112 ALFFPYRTMGX--KEQDGGYYCVAKNVRQGAVSRRASLQIAVLRDDFRVEPKDTRVAKG 169
   :| : || |:| |:| | |:| | |:|:| |:| |:| |:| : |
Db 88 SLFLLVKNSGKNGKSDAGAYVCASNEHEGVKSNESGLKAMLREDFRVRPRTVQALGG 147

Qy 170 ETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFAGSSRRVIRDGNNLLISNVEPIDEG 229
   |:| |:| |:| |:| : || |:| | : |:| |:| |:| |
Db 148 EMAVLECSPPRGPFVPPVSWRKDD-----KELRIQDMPRTYHLSDGNLIIDPVDRSDSG 201

Qy 230 NYKCIAQNLVSTRESSYAKLIVQVKPYFMKEPKDQVMLYGQTATFHCSVGGDPPPKVLWK 289
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 202 TYQCANNMVGERSVNPARLSVFEXPKFEQPKDMTVDVGAALVDFCRVTGDPQQTW 261

Qy 290 KEEGNIPVSRARILHDEKSLEISNITPTDEGTIVCEAHNNVQGISARASLIVHAPPNFTK 349
   :: :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 262 RKNPEMPVTRAYIAKDNRGLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPPSFQT 321

Qy 350 RPSNKKVGLNGLVQLPCMASGNPPPSVFWTKEGVSTLMFPN--SSHGRQYVAADGTLQIT 407
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 322 KPADQSVPAGTATPECTLVGQSPAYFWSKEGQDQLFPYSVSADGRTKVSPTGTLTIE 381

Qy 408 DVRQEDEGGYYVCSAFSVSDSSTVRVLQVS-----SVDERPPPIIQIGPANQTLPGKSV 461
   |:| |:| |:| |:| : || |:| : |:| |:| |:| |:| |:|
Db 382 EYRVDEGAYCAGNNSAGSSLSKAALKVTTTAVTNGTTPAKPPPTIEGHONOTLMVGSS 441

```

```

RESULT      2
T30805
dutt1 protein - mouse
N:Alternate names: transmembrane receptor protein Robo1 homolog
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30805

```

```

Qy 56 PRIEHPTDLVVKKNPATLNCKVEGKPEPTIEWKDGPEVST--NEKSHRVQFKDGL 113
   |||:|||:|||:| ||||| |||'| |||||:| || | | :: :|||: | :|
Db 29 PRIVEHPSDLIVSKGEPATLNCKAERGPTPTIEWYKGGERVETDKDPRSHRMLPSGSL 88

Qy 114 FFYRTMQGKKFQ-DGGEYWCVAKNRVGOAVSRHASLQIAVLRDDFRVEPKDTRVAKGETA 172
   || | : | :| : | | |||:| :|:| | :|||:| |||| | | | | | |
Db 89 FFLRIVHGKRSRPDEGVYICVARNYLGEAVSHNASLEVAILRDDFRQNPSDVMVAVGEPA 148

Qy 173 LLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVIVDGGNLLISNVEPIDEGNYK 232
   :|| |||:| ||||:| ||| |||| | : | || |:|: | | |
Db 149 VMECQPPRGHPEPTISWKKDGSP added-----KDERITI-RGGKLIMITYRKSDAGKYV 200

Qy 233 CIAQNLVGTRESSYAKLIVQVKPYFMKEPKDQVMLYGQTATFHCSVGGDPPPKVLWKKEE 292
   |: |:| ||| |:| | :| |:| : : |:| | ||| | |:|:|
Db 201 CVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTDDSAEFKCEARGDPVPTVRWKDD 260

Qy 293 GNIPVSRARILHDEKSLAISNITPDEGTVCYCAHNNVGOISARASLIVHAPPNFTKRPS 352
   |:| | |:|:| |:| | |:| | | |:| |:| |:| |:| |:|
Db 261 GELPKSRYEI-RDDHTLKIRKVTAGDMGSYTCVAENVMVGKAEASATLVQEPHPHFVKPR 319

Qy 353 NKKVGLNGVVQLPCMASGNPPPSVFTWKEGVSTL added---PNSSHGRQYVAADGTLQITDV 409
   :: | | | :||| |:|:| |:| |:| | | | | | |:|:|
Db 320 DQVVALGRTVTFQCEATGNPQPAIFWRREGSQNLLFSYQPPQSSSRFSVSQTGDLTITNV 379

Qy 410 RQDEGEYYVCSAFSVSDSSTVRVFLQVSSV-DERPPPIIQIGPANQTLPGKSVAILPCRA 468
   :: | |||:| :| | : |:|:| | :|||:|: || |||: | | |
Db 380 QRSDVGYIQCITLNVAGSIITKAYLEVTDVIADRPVPIRQGPVQNTVAVDGTLLSCVA 439

Qy 469 TGNPSPRIKWFHDHGAQVA-GNRYSIIQGSSLRVDDLQLSDSGTYTCTASGERGETSWAA 527
   |||:| | | | | | | | | | | | | | | | | | | | | |
Db 440 TGSPAPTILWAKDGLVSTQDSRIKQLESGVLQIRYAKLGDTGRYCTASTPSGEATWSA 499

Qy 528 TLTVKEPG-STSLHRAADPSTYPAPPGTPKVLNVSRSTSISLRWAKSQEKGAVGPIIGYT 586
   |:|:| | | | | | | | | | | | | | | | | | | | |
Db 500 YIEVQFEGVYPQPPRPDPNLIPSAFSPKPEVTDVSKNTVTL added---QPNLNSGATPTSYI 556

Qy 587 VEYFSPDLQTCWIVAHRVGDQTVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKI 646
   :| | | :| | | | | | | | | | | | | | | | | | | |
Db 557 IEAFSHASGSSWQTAENKVTETFAIKGLKPNAILYLFLVRAANAYGISDPSQISDPVKQT 616

Qy 647 EADFDAASANDLSAARTLLTGKSV-ELIDASAINASAVRLEWMLHVSADSEKYEGLRIHY 705
   : : : : : : : : : : : : : : : : : : : : : :
Db 617 DVPPTSQGVDDHKVQREL--GNVVLHLNPTILSSSSVEVHWT--VDQSQYIQYKILY 672

Qy 706 K--DASVPSAQYHSITVMDASAESFVGNLKKYTYEFFLTPFFETIEIQGPSNKTALTY 763
   : : : : : : : : : : : : : : : : : : : : :
Db 673 RPSGASHGESEWLVEVRTPTKNSVVIDPLRKGVNIEIKARPPFNEFGADSEIKFAKTL 732

Qy 764 EDVPSAPPDNIQIGMY--NQTAGVVRWTPPPSQHHNGNLYGKIEVSAGNTMKVLANMIL 821
   |:| ||||:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 733 EEAPSAPPRSVTYSKNDGNGTALIVTWQPPPEDTQNGMVQEYKV-WCLGNETKYHINKTV 791

Qy 822 NATTISVLLNLLTGAVISVRLNSFTKAGDGPISKPISLFMDPTHVHPPRAHPSGTHDG 881
   : : : : : : : : : : : : : : : : : : : : :

```

```

Qy      44  NGLPA-----VRGQYQSPRIIEHTDLVVKKNEPATLNCKVEGKPEPTI  87
      || ||                      : | : |||:|||:||| | ||||| ||| ||
Db      40  NGTPAPTSDDNDNSLGYTGSRRLQEDFPFRIVEHPSDLIVSGEPATLNCKAEGRPITPI  99

Qy      88  EWFKDGEPVST--NEKKSHRVQFKDGLFFYRTMGKKEQ-DGGEYWCVAKNRVQGAVER  144
      ||:| || | | :: :||| :||| : : :| : | |||:| :|:||
Db     100  EWYKGGERVETDKDDPRSHRMLLPGLSGSLFFLRIVHGRKSRPDEGVYICVARNYLGEAVSH  159

Qy     145  HASLQIAVLRRDDFRVEPKDTRVAKGETALLECGPKGIPEPTLIWIKDGVPLDDLKAMSF  204
      :|||:|||:|||:||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

Db 160 NASLEVAILLRDFRQNPSDVMVAVGEPAYMECQPPRGHPHEPTISWKKDGSPLDD----- 213
Qy 205 GASSRVRIVDGGNLLISNVPEIDEGNKYCIAQNLVGTRESSYAKLIVQVKPYFMKEPKDQ 264
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 214 -KDERITI-RGGKLMITYTRKSDAGKYVCVTNMVGERESKVADVTLERPSFVKRPSNL 271
Qy 265 VMLYGTATFHCSSVGGDPPPKVLWKKEGNIPVSRARILHDEKSLEISNITPTDEGTIVC 324
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 272 AVTVDDSAEFKCEARGDPVPTFGWRKDDGELPKSRYEI-RDDHTLKIRKVITAGDMGSYTC 330
Qy 325 EAHNNVGQISARASLIVHAPPNFTKRPSNKKVGLNGVQLPCMASGNPPPSVFWTKEGVS 384
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 331 VAENMVGAKEASATLTVQEPHFVVKPRDQVVALGRVTVTFQCEATGNPQPAIFWRREGSQ 390
Qy 385 TLMF---PNSHGRQYVAADGTLQITDVRQEDEGYIVCSAFSVSDSTVRVFLQVSSV-D 440
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 391 NLLFSYQPPQSSSRFSVSQIGDLTVTNVQRSDVGYYICQTLNVAGSIITKAYLEVTDVIA 450
Qy 441 ERPPPIIQIGPANQTLPKGSVATLPCRATGNPSPRIKWFHDGHAVQA-GNRYSIQGSGL 499
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 451 DRPPPVIRQGPVQNTVAVDGTLTLCSCVATGSPVPTILWRKDGLVSTQDSRIKQLESGLV 510
Qy 500 RVDDLQLSDSGTYTCTASGERGETSWAATLTVEKPG-STSLHRAADPSTYPAPPGTPKVL 558
   :: |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 511 QIRYAKLGDTSRYTCTASTPSGEATWSAYIEVQEFQVPVQPPRTPDNLIPSAPSKPEVT 570
Qy 559 NVSRTSISLRWAKSQEKPGAVGPIIGTYVEYFSPDLQGWIVAHRVGDQTQVTSGLTPG 618
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 571 DVSKNVTLLW---QPNLNSGATPTSIIIEAFSHAGSSWQTVAENVKTTETFAIKGLKPN 627
Qy 619 TSYVFLVRAENTQGISVPSGLSNVIKTIADFAASANDLSAARTLLTGKSV-ELIDASA 677
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 628 AIYLFVLVRAANAYGISDPSQISDPVKTDQVPPTTQGVDRHQVQREL--GNVVLHLHNPTI 685
Qy 678 INASAVRLEWLHLSVADEKYVEGLRIHYK--DASVPSAQHYSITVMDASAESFVGNLKK 735
   :::: |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 686 LSSSSVEVHWI--VDQQSQYIQGYKILYRPSGASHGESEWLFVVRTPTKNSVVIPDLRK 743
Qy 736 YTKYEFLTPPFETIEGQPSNKALTIEDVPSAPPDNIQIGMY--NQTAGWVRWTPPPS 793
   || |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 744 GVNYEIKARPPFNEFQGDSEIKFAKTLEERPSAPPSPVTSKNDGNGTAILVTWQPPE 803
Qy 794 QHHNGNLYGYKIEVSAGNTMKVLANMTLNATTSVLLNNLTGAVYSVRLNSFTKAGDGP 853
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 804 DTQNGMVQEYKV-WCLGNETRTHINKTVDGSTFSVVIFFLVPGIRYSVEVAASTAGPGV 862
Qy 854 YSKPISLFMDPTHHVHPRAHPSGTHDRHEGQDLTYHNNGN-IPPGDINPTTHKKTDTY 912
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 863 KSEPQFIQLD-----SHGNPVSPED-QVSLAQQISDV 893
Qy 913 LSGP-----WLMVLVCIVLLVLVISAAISMVYFKRKHQ----- 945
   : |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 894 VKQPAFIAGIGAACWIIWMFSIWL-----YRHKRRKRLSGSYAGIRKVPST 942
Qy 946 ---MTKELGHLVSVDNEITALNINSKESL-WIDHHRGW-RTADTDK-----SG 990
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 943 FFTPVTYQRGGEAVSSGGRPGLLNISEPATQPWLAD--TWPNTGNSHNDCSINCTASNG 1000
Qy 991 LSESKLLSHVNSSQ--SNYNNS-----DGGTDYAEVDTRNLTTFFYNCRKSPD--- 1035
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 1001 NSDSNLTYTYSRPADCIANYNLQDNKQTNLMLPESTYGVGDVLSKNKINEMKTFNSPNLKD 1060
Qy 1036 -----NPTPYATTMII-----GTSSSE----- 1052
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 1061 GRFVNPSPGQPTPYATTQLIQANLINNMNNGGDSSEKHKWPQGQKQGEVAPIQYNIMEQN 1120
Qy 1053 -----TCIKTTSISADKDSGTHSPYSDAFAGQVPAVPVVKSNYLQYPVEP---I 1098
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 1121 KLNKDYRANDTIILPTIPYNHSDYQNTGGSYNSDDRGSSSTSGSQGHKKGARTPKAPQCGM 1180
Qy 1099 NWSEFLPPPPPEHPPPSST-----YGYAQGSP-----ESSRKSSK 1132
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 1181 NWADLLPPPAHPHPHSNSEEYSMSVDESVDQEMPCVPVPMARYLQODELEEEEAERGPT 1240

```


RESULT 5
T29549
hypothetical protein ZK377.3 - *Caenorhabditis elegans*

```

RESULT 6
T29548
hypothetical protein ZK377.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29548
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK377.
A:Reference number: Z20639
A:Accession: T29548
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-874 <NHA>
A:Cross-references: EMBL:U88183; PIDN:AAE52657.1; GSPDB:GN00028; CESP:ZK377.2
A:Experimental source: strain Bristol N2; clone ZK377
C:Genetics:
A:Gene: CESP:ZK377.2
A:Map position: X
A:Introns: 91/2; 356/1; 452/1; 701/3; 746/3; 850/1

```

[illegible]

RESULT 7
T30532
neural cell adhesion molecule L1 homolog - Fugu rubripes

Query Match 9.1%; Score 677; DB 2; Length 1443;
Best Local Similarity 23.1%; Pred. No. 1.4e-26;
Matches 304; Conservative 171; Mismatches 541; Indels 302; Gaps 49;

Qy 157 FRVEPKDTRVAKGETALLECG----PPKGIPEPTLIWIKDGVPDLDKAMSFSGASSVR 211
| | | | : | : : | ||| : | ||| : : | :
Db 21 FLVEPMDILSVRGASVMNCSSYCETPPK-----IEWKKDGT-----LLNLVSDDRRQ 68

Qy 212 IVDGGNLLISNV-----EPIDEGNYKCIAQ-NLVGTRRESSYAKLIVQVKPYFMKEPKDQ 264
: : : : : : : : : : : : : : : : : :

```
Qy 1257 PSPMQP-PPPVVPEGWYQPVHPNSHPMHTSSNHQIQCSSE-----CS 1300
      | : | | : | : | | : | : | : | : | : | : | : | : | : |
Db 1060 PVDVGDPKPKPLSGSNS-----PHGSPTSPLDNSMLLVIIYVSGVITIVIVIVAVFCT 1113

Qy 1301 DHSRSSQSHKRLQLLEEHGSSAKQGG-----HHRH-----RAPVVOPCM 1340
      | : | | | : | : | : | : | : | : | : | : | : | : | : |
Db 1114 RRTTSQHKKRAACKSVNGSH-KYKGNKSDVKPPDLWIHERLELKPDKSPDPNPIM 1170
```

RESULT 9
 T08851
 Down syndrome cell adhesion protein 1 - human (fragment)
 N;Alternate names: Down syndrome cell adhesion molecule
 C;Species: Homo sapiens (man)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
 C;Accession: T08851
 R;Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenb
 submitted to the EMBL Data Library, September 1997
 A;Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down
 A;Reference number: 216495
 A;Accession: T08851
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1896 <YAM>
 A;Cross-references: EMBL:AF023449; NID:g3169765; PID:g3169766
 A;Experimental source: brain; developmental stage: 14 weeks; fetal
 C;Genetics:
 A;Gene: DSCAM
 A;Map position: 21q22
 A;Note: derived from alternately-spliced mRNA
 C;Function:
 A;Description: involved in nervous system development
 C;Keywords: alternative splicing

Query Match 8.9%; Score 661; DB 2; Length 1896;
Best Local Similarity 24.2%; Pred. No. 1.2e-25;
Matches 305; Conservative 175; Mismatches 473; Indels 308; Gaps 60;

```

Qy      55 SPRIIEHPDLVVKKNEPATLNCXVEGKPEPTIEWFKDGEPVSTNEKKSHRVQ---FKDG 111
      :|:| | : :| | | :| :| | | | | :| :| : :| | :|
Db      392 TPKIISAFSEKVVSPAEPVSLCMNVKGTPLPTITWTLDDDPIL---KGGSHRISQMITSEG 449

Qy      112 ALFFYRTMQGKKEQDGGEYWCVAKNRVQAVSRHASLQIAVLRDDFRVEP-KDTRVAKGE 170
      :| :| : :| | | | | | :| :| : :| :| :|
Db      450 NVVSYLNISSQVRDGGVRYCTANNSAG-VVLYQARINV---RGPASIRPMKNITAIAGR 505

Qy      171 TALLECGPPKPIGEPTLWIWKDG---VPLDDLKAMSGASSRVIVDGGNLLISNVE-PID 227
      :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      506 DTYIHC-RVIGYPYISIKWYKNSNLLPFN-----HRQVAFENNNGTLKLSDVQKEVD 555

Qy      228 EGNKYC---IAQNLVGTRESSYAKLIVQVKPYF--MKEPKDQVMLYIGQTATFHC-SVGGDP 282
      || | :| :| :| :| :| :| :| :| :| :| :|
Db      556 EGEYTCNVLVQQLSTSQSVH---VTVKVPPIQFEPFRFSI---QQRVFIPCVVVSGDL 610

Qy      283 PPKVLWKKKEGNIPIVSRARILHD---EKSLEISNITPDEGTYVCEAHNNVQGISARASL 339
      | :| :| :| | :| :| :| | | | | | :| :| :|
Db      611 PITITWQGRPIPGSLGVTIDNIDFTSSLRISNLSLMHNGNYTCIARNEAAAEHQSQL 670

Qy      340 IVHAPPNTFRPSNKKVGLNG-VVQLPCMAGSNPPPVFWT-KEGVSTLMF-PNSSHGRQ 396
      || || | :| :| :| | | | | | :| :| :| :|
Db      671 IVRVPPKFVVQPRDQ-GIYGKAVILNCSABGYVPVPIVWKFSKAGVPOFQPIALNGRI 729

Qy      397 TYAADGTQLTIDVRQEDGEGYVCSAFSVSDSSTVR-VFLQVSSVDERPPPIIQIGPANQT 455
      :| :| :| :| :| | | | :| :| : :| | :| :| :|
Db      730 QVLSNGSLLIKHVVEEDSGYILCKVSNVDGADYSKSMYLTV-----KIPAMITSYPNTTL 784

Qy      456 LPKGSVATLPCRAATGNPSPRIKWFHDGHAVQAG-NRYSIIQG-----SSLRVDDLQLS 507
      :| :| :| | :| :| :| :| :| :| :| :| :|
Db      785 ATQGQKKEKMSCTARGEKPIIVRWKEKDRIINPEMAYRLVSTKEVGEEVISTLQILPTVRE 844

Qy      508 DSGTYICTASGERGETSWAATLTVKEKPGSTSLHRAADPSTYPAPPGTKPVLNVSRSTSISL 567

```

Db 845 DSGFFSCHAINS YGEDRGIIQLTVQEP-----PDPPEI-EIKDVKARTITL 889
Qy 568 RWAQSEKPGAVGPIIGTYVEYFSPDLQGWIVAAHRVGD-----TQVTISGLTPGTSYV 622
Db 890 RWTMGFD---GNSPITGYDIE--CKNKSDSW-DSAQRTKDVSPQLNSATIIDIHPSSYTS 943
Qy 623 FLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSAARTLLTGKSVELIDASAINASA 682
Db 944 IRMYAKNRIGKSEP---SNEL-TITAD-EAAP-----DGPPQE-VHLEPISSQS 986
Qy 683 VRLEWMLHVSADKEY----VEGLRIHYKDVSAQYHSITVMDASAES--FVVGNLKK 735
Db 987 IRVTW---KAPKKHLQNGIIRGTYQIGREYSTGGNFQFNISVDTSGDSEYVITDLNLMK 1042
Qy 736 YTKYEFPLTPFFETIEGQPSNSKTALTYEDVPSAPPDNIQIGMNQTAGWVRWTPPPSQH 795
Db 1043 FTQYGLVQACNRAGTGPPSSQEIITTTLEDVPSYPPENVQAIATSPESISISWSTLSKEA 1102
Qy 796 HNGNLYGYKIEVSAGNTMKVLANMTLNATTT--SVLLNNLTGAVYSVRLNSFTKAGDGP 853
Db 1103 LMGILQGRFV-IYWANLMDGELGEIKNITTTPQSLDGLKTYNYSIQVLAFTAGDGV 1161
Qy 854 YSKPISLMDPPTHVHPRAHPSGTHDGRHEGDLTYHNNGNIPPGDINPTTHKKTIDYL 913
Db 1162 RSEQI--PTRTKEDVPGP---PAG----- 1180
Qy 914 SGPWLMVLCIVLLVLSAISMVYFK-----RKQMTKELGHLVSVSDNE-- 960
Db 1181 -----VKAAASASMVFSWLPPLKNGIIRKYTVFCSHPYTPVISSEFAS 1226
Qy 961 -----ITALNINSKESLWIDHHRGWRTADTDKDSLSESKLLSHVNSSQSNYNNSDGG 1013
Db 1227 PDSFSYRIPNLSRNQYSVWV-----VAVTSAGRG-----NSSEII 1262
Qy 1014 T--DYAEVDTRNLT-----TFYNCRKSPDNPTPYATTM--IIGTSSETCTKT 1057
Db 1263 TVEPLAKAPARILTFSGVTTPWMDKIDVLC-KAVGDPSPAVKWMKDSNGTSPSLVTIDGR 1321
Qy 1058 TSISAD-----KDSGTHSPYSDAFAGQVAPVVKSNYLYQVPEPINWSE---- 1102
Db 1322 RSIPFNGSFIIRTVKAEDSGYYS-----CIANN-----NWSDEIIL 1358
Qy 1103 --FLPPPPPEHP-----PPSSYTYAQSGPESSRKSSKSGISTNQSIILNASHSS 1152
Db 1359 NLQVQVPPDQRLTVSKTSSSITLSWLPD-----NGGSSIRGYILQYSEDNS 1407
Qy 1153 SSGGFSAMGVSQYAVACPPENVISNPLSAGAGTQNRQITPTNQHPQLPAYFATTGP 1212
Db 1408 EQ-----WGSFP---ISPERSYR--LENLKCCTWYKTLTAQN-----GVGP 1445
Qy 1213 G 1213
Db 1446 G 1446

RESULT 10
A34695

axonal glycoprotein TAG-1 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 21-Jan-2000

C:Accession: A34695

R:Furley, A.J.; Morton, S.B.; Manalo, D.; Karagogeos, D.; Dodd, J.; Jessell, T.M.
Cell 61, 157-170, 1990

A:Title: The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neur

A:Reference number: A34695; MUID:90199890

A:Accession: A34695

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1040 <P>

A:Cross-references: GB:M31725; NID:g207148; PIDN:AAA2201.1; PID:g207149

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

C:Keywords: glycoprotein

F;343-399/Domain: immunoglobulin homology <IMM>

Query Match 8.9%; Score 658; DB 2; Length 1040;
Best Local Similarity 24.1%; Pred. No. 8.1e-26;
Matches 251; Conservative 145; Mismatches 433; Indels 214; Gaps 36;

Qy 22 SRSRSSRMWLLPAWLLLVLA---SNGLPVARGQYQSPRIIEHPTDLVVK---NEPAT 74
Db 5 ARKKASLLLVLATVALVSSPGWSFAQGTPTATG---PIFEEQIGLLFPEESAEDQVT 60
Qy 75 LNCKVEGKPEPTIEWFKDGEVPVSTNEKSHRVQFKDGFYRTMQGKEQDGGYWCVA 134
Db 61 LACRARASPPATYRWKMG--TDMNLEPGSRHQLMGNLVI---MSPTKQDAGVYQCLA 115
Qy 135 KNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKETALLECGPPKIGPEPTLIWI---- 190
Db 116 SNPVGTVSKAEVLRFGFLQEFSEKEDRPVKTHEGWGMPLCNPPAHYGLSYRLLNEF 175
Qy 191 KDGVPDDLKAMSFGASSRRVIRVDGNNLISNVEPIDEGNYKCIANLV--GTRE--SSY 245
Db 176 PNFIPDGRHFVS-----QTTGNLYIARTNASDLGNYSCLATSHMDFSTKSVFSKF 226
Qy 247 AKLIV-----QVKPYFMKEPKDQVMLYQGTATFHCSVGGDPPKVLWKKEGNIP 296
Db 227 AQLNLAEDPRLFAPSIKARF---PPETYALVGQVTLCEFAFGNVPVRKWKVDSGLS 283
Qy 297 VSRARILHDEKSLSEISNITPTDEGTYVCEAHNNVQGISARASLIVHAPPNFTKPSNKKV 356
Db 284 PQWATA---EPTLQIPSVSFEDEGTYECEANSKGRDVTQGRIVVQAPQEWLVISDTFA 340
Qy 357 GLNGVVQLPCMASGNPPSPVFTKEGVSTLMFPNSHGRQYVAADGTLQITDVRQDEGY 416
Db 341 DIGSNLRWCCAAAGKPRPVRWLNGE-----PLASQNRVEVLA-GDLRFSLKSLSDSGM 394
Qy 417 YVCASFVSVD--STVRVFLQVSSVDERPPPIIIGPANQTLPKGSVATLPCRATGNPSP 474
Db 395 YQCAENKHTIYASAEALVQALAPDFRQNPVRRILPA---ARGEISILCPRAAPKA 450
Qy 475 RIKWFHDGHAVQAGNRYSIQGSRLVDDLQLSDSGTYCTTASGERGETSWAATLTVE-- 532
Db 451 TILWSKGTIELGNSTRVTVTSDGLIIRNISRSDGKYTCFAENFMKANSTGILSVRDA 510
Qy 533 -----KPGSTSL-----HRAADPS---TY----- 548
Db 511 TKITLAPASADINVDNLTQCHASDHPTMDLFTFTWLDLDFIDFKPGGHYRRASAKET 570
Qy 549 -----PAPPPTPKVLNVSRISLRLW 569
Db 571 IGDLTILNAHVHRGGKYTCMAQTVVDGTSKEATVLVLRGPPGPGVVRIDGTVQLSW 630
Qy 570 AKSQEKPGAVGPIIGTYVEYFSPDLQGWIVAAHRV---GDTQVT--ISGLTPGTSYVFLV 625
Db 631 SRGPDN---SPIAKYTLQARTPPSGKWKQVTRNPVNIENGAETAQVLGLMPMDYEFVR 687
Qy 626 RAENTQGISVPSGLSNVKTIEADFDAASANDLSAARTLLTGKSVELIDASAINASAVRL 685
Db 688 SASNLTGTEPSPGSSKIRTKEA-VPSVAPSGLSGG---GGAPGELI-----I 731
Qy 686 EWLHVSADKEYVEGLRIHYKDVSAQYHSITVMDASAESFVVGN--LKKYTYEFL 743
Db 732 NW-TPVSRBYQNGDGFYLLSFRQGSQWQARTVPGADAQYFYVGNDSIQPTPFVEKI 790
Qy 744 TPFETIEGQPSNSKTALTY--EDVPSAPPDNIQIGMNQTAGWVRWTPPPSQHNGNLY 801
Db 791 RSY--NRRGDGPESLTALVISAEEPRVAPAKVWAKSSSEMNVSW-EPVLQDMNGILL 847
Qy 802 GYKIEV-SAGNTMKVLANMTLNATTSVLLNNLTGAVYSVRLNSFTKAGDGPYKPSIL 860
Db 848 GYEIRYWKAGDNEAAADRVRTAGLTSARVTGLNPNTKYHVTVRAYNRAGTGASPSADA 907
Qy 861 FMDPPTHVHPRAHPSGT-----HGRHEGQDLTYHNNGNIPPG 899

Qy 310 EISNITPTDEGYVCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVLGNGVVQLPCMAS 369
||| || | |: || :| :: :: :: :: ||
Db 283 ETONEFOEDTGSYEGIAEHSRGKNVARGRIITYAKPYWLOLRDEIJAVEDSIIWEFCRAS 342

Qv 695 E-KYVEGLR----IHYKDASVPSAQYHSITVMDASAESFVVGN--LKKYTKYEFLTPPF 747

Db 743 GFGYLLSFRRGSTHWQARVPG-----ADAQYFVYSNESVRPIPEVKIRSY- 791
Qy 748 ETIEGQPSNSKTALTY--EDVPSAPPDNIQIGMYNOTAGVVRWTPPPSQHHNGLYGI 805
Db 792 -NRRGDGPESLTALVYSABEEPRVAPTQVWAGVSSSEMNTW-EPVQDMGILLGYEI 849
Qy 806 EV-SAGTMKVLNMTLNTATTSVLLNLTGAVYSVRLNSFTKAGDGPYSKPI-SLFMD 863
Db 850 RYWKAGDKEAAADRVRTAGLDTARSVGLHPNTKYHTVTRAYNRACTGPASPSANATMMK 909
Qy 864 PTHHVHPRAHPSG-----THDGRHEGQDLTYHNNGNIPPGDIN 903
Db 910 P-----PRRPPGNISWTFSSSLSIKWDPVVPFERNESAVTGKMLYQNDLH-----LTP 959
Qy 904 TTHKKTDTYLSGP 916
Db 960 TLHLTGKNWIEIP 972

RESULT 14
S26180
neurofascin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jan-2000
C:Accession: S26180
R:Volkmer, H.; Hassel, B.; Wolff, J.M.; Frank, R.; Rathjen, F.G.
J. Cell Biol. 118, 149-161, 1992
A:Title: Structure of the axonal surface recognition molecule neurofascin and its relat
A:Reference number: S26180; MUID:92317154
A:Accession: S26180
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1272 <VOL>
A:Cross-references: EMBL:X65224; NID:g63659; PIDN:CAA46330.1; PID:g63660
C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; 1
F:279-336/Domain: immunoglobulin homology <IMM>

Query Match 8.7%; Score 644; DB 2; Length 1272;
Best Local Similarity 21.6%; Pred. No. 5.4e-25;
Matches 272; Conservative 166; Mismatches 455; Indels 368; Gaps 39;

Qy 54 QSPRIIEHT-DLVVKNKNEPATLNCKVEGKPEPTIEFWKDGEPVSTNEKSHRVQFKDGA 112
Db 40 QPPTITKQSVKDIYDPRDNIFIECEAKGNPVPPTFSWRNKGKFFNVAKDPKVMRRRSGT 99
113 LFFYRTMGKKQDQGEYWCVAKNRVGQVSRHASLQIAVLRDDFRVEPKD----TRVAK 168
Db 100 LVIDFHGGGRPDYEGEYQCFARNDYGTALSSKHLQVS----RSLWPKEKVDVIEVDE 155
Qy 169 GETALLECGPPKGIPEPTLIWIKDGV-PLDDLKAMSGASSRVIVDGGNLLISNVEPID 227
Db 156 GAPLSLQCNPPGLPPPVPFWMSSMEPIHQDKRVSG-----QNGDLTFSNVMLQD 207
Qy 228 -EGNYKCIQ-----NLVGTRESSYAKLIVQVKPYFM-- 258
Db 208 AQTDYSCNARFHTHTIQQKNPYTLKVKTKKPHNETSLRNHTDMISARGVTETTPSFMY 267
Qy 259 -KEPKDQVMYLGQTATFHCSVGGDPPPKVLWKEEGNIPVSRARILHDEKSLEISNITPT 317
Db 268 YGTSSSQMVLRGVLDLLECIASGVAPADIMWYKGGELPAGTKLENFNKALRISNVSEE 327
Qy 318 DEGTYYCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVGLNGVQLPCMASGNPPPSVF 377
Db 328 DSGEYFCLASNMKSIRHTISVRVKAAPYWLDEPQNLILAPGEDGRLVCRANGNPKPSIQ 387
Qy 378 W-----TKEGVSTLMFPNSSHGRQYVAA----- 400
Db 388 WLNVGEPIEGSPNPSPREVAGDTIVFRDTQIGSSAVYQCASNEHGYLLANAFVSLDVP 447
Qy 401 -----DGTILQIT 407

Db 448 PRILAPRNQLIKVIQYNRTRLDCCPFPGSPIPLRWFKNGQGNMDDGGNYKAHENGSLMS 507
Qy 408 DVRQDEGEYVCSAFSVVDSSTVRVLQVSSVDERPPPIIQIGPANQTLPGKSVATLPCR 467
Db 508 MARKEDQGIYTCVATNIGKVEAQRVLEV-----KDPTIRVIRGPDQVVRGSMPLHCR 562
Qy 468 ATGNPSPR--IKWFHDGHAVQAGNRYIIQSSSLRVDDLQSLDSGTYTCTASGERGETSW 525
Db 563 VKHDTPLKLTITWTKDDAPLYIGNRMK-KEDDGLTIYGAEKDQGDYTCVATSLDQDSA 621
Qy 526 AATLVEKPGSTSLHRAAD-PSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIG 584
Db 622 KAYLTVL---AIPANRLDLPKERPDRDRLESLDLAERSVKLTWIPGDNN---SPITD 675
Qy 585 YTYEYFSPDLQTC-WIVAHRVGDQTQVTSGLTPTGTSYVFLVRAENTQGISVPSGLSNVI 643
Db 676 YIVQFEEDRFQGTWHNHSRYPGVNSALLSLSPVINYQFVIAVNDVGSLSFSPMPERY 735
Qy 644 KTIEADFDASANDLSAARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLR- 702
Db 736 QT-----SGARPEINPTGVQ---GAGTQKNMEITW-TPLNATQAYGNLRY 778
Qy 703 -IHKDASVPSAQYHSITVMDASAEFVVGNLKXYTKEFFLTPFFETIEGQPSNSKTAL 761
Db 779 IYVRRRD-PRGSWYNETV---KAPRHVVWNTPIYVYIKVQA--ENDFGRAPEPTIYI 832
Qy 762 TY--EDVPSAPPDNIQIGMYNOTAGVVRWTP----- 789
Db 833 GYSGEDYPKAAPDVRIRVNLSTALITWTRVHLDTIQQLKEYRAYFWRDSSLKLNW 892
Qy 790 -----PPPS----- 793
Db 893 SKKQYVSPFGDRNRGIVSRFPYSNYKLEMVVTNGRGDGRSEVKEFPTPEGVSPSPRY 952
Qy 794 -----QHNGNLYGYKIEVSAGNTMK---VLNMTLNTATTSVLLN 831
Db 953 LRIRQPNLESINLEWDHPEHNGVLTYGNLRYQAFNGSKTGRTLVENFSNPQTRFTVQRT 1012
Qy 832 NLTTGAVYSVRLNSFTKAGDGPY-----SKPISLFMDPTHHVHPRAHPSGT 878
Db 1013 DPISR--YRFFLRARTQVGDGEVIVEESPALLNEATPTPASTWLPPPTTELTPAATIATT 1070
Qy 879 HDGRHEGQDLTYHNNGNIPPGDIN---PTTHKKT----- 910
Db 1071 -----TTTATPTTETPTEIPTTAIPTTTTTATAASTVSTTTTAAAAAATTKQ 1122
Qy 911 -----DYLGGPWLMLVC-IVLLVLVISAISMVYFKRKHQMTKELGHSVVDNE 960
Db 1123 ELAYTKNHVDIATQGWFIGLMCAIALLLVIL---LIVCFIKRSR-----GGKYPVRDNK 1173
Qy 961 ITALNINSKESLWIDHHRGWRTADTKDS---GLSESKLLSHVNSQSNYNNSDGGTDYA 1017
Db 1174 DEHLNPEDKNV--EDGSFYRSLESDENKPLNPQTSLDGTIKQESD-----DSLVDYG 1227
Qy 1018 E 1018
Db 1228 E 1228

RESULT 15
S05479
neural cell adhesion molecule L1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S05479; B60850; S22167
R:Moos, M.; Tacke, R.; Scherer, H.; Teplow, D.; Frueh, K.; Schachner, M.
Nature 334, 701-703, 1988
A:Title: Neural adhesion molecule L1 as a member of the immunoglobulin superfamily wi
A:Reference number: S05479; MUID:88318924
A:Accession: S05479
A:Molecule type: mRNA
A:Residues: 1-1260 <MOO>
A:Cross-references: EMBL:X12875; NID:g53336; PIDN:CAA1368.1; PID:g53337

A;Note: the authors translated the codon CCT for residue 166 as Leu, ACT for residue 396
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Rathjen, F.G.; Wolff, J.M.; Frank, R.; Bonhoeffer, F.; Rutishauser, U.
J. Cell Biol. 104, 343-353, 1987
A;Title: Membrane glycoproteins involved in neurite fasciculation.
A;Reference number: A60850; MUID:87109457
A;Accession: B60850
A;Molecule type: protein
A;Residues: 20-28,'XX',31-36 <RAT>
R;Kohl, A.; Giese, K.P.; Mohajeri, M.H.; Montag, D.; Moos, M.; Schachner, M.
submitted to the EMBL Data Library, December 1991
A;Description: Analysis of promoter activity and 5' genomic structure of the neural cell
A;Reference number: S22167
A;Accession: S22167
A;Molecule type: DNA
A;Residues: 1-165,'L',167-189,'E',191-281,'S',283-395,'S',397-514,'APEKNPVDV',524,'GEGNE
A;Cross-references: EMBL:X63511
C;Genetics:
A;Introns: 26/1; 31/1; 66/2; 133/1; 174/1; 231/1; 268/2; 330/1; 374/1; 422/1; 459/2
A;Note: the list of introns may be incomplete
C;Superfamily: neural cell adhesion molecule LI; fibronectin type III repeat homology; 1
Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmembran
1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: neural cell adhesion molecule #status experimental <MAT>
F;256-313/Domain: immunoglobulin homology <IMM1>
F;440-498/Domain: immunoglobulin homology <IMM2>
F;531-592/Domain: immunoglobulin homology <IMM3>

Query Match 8.6%; Score 641.5; DB 1; Length 1260;
Best Local Similarity 23.0%; Pred. No. 7.1e-25;
Matches 275; Conservative 167; Mismatches 458; Indels 295; Gaps 48;

Qy 35 WLLLVLASNGLPAVRGQYQSPRIIE-----HPTDLVVKNEPATLNCKVEGKPEPTI 87
| | | : | : : : | | | : | : : :
Db 9 WPLL-LCSPCLLIQIPDEYKHHVLEPPVITEQSPRLVVPPTDDISLCKEARGRQVEF 67

Qy 88 ENFKDGEVPSTNEKKS---HRVQPKDGLFFYRTMGKK---EQDGGEYVCAKNRVGQA 141
| | | : | : : : | : | : : | | | : : |
Db 68 RWTKDGIHFKPKKEELGVVHEAPY-SGSF-----TIEGNNFAQRFGIYRCYASNKLGA 122

Qy 142 VSRHASLQIAVLRDDFRVEPKDT---RVAKGETALLECGPKGIPEPTLIWIKDGV--- 194
: | : : : : | : | : : | | : : : :
Db 123 MSH----EIQLVAEGAPKWPKETVKEVEGESVVLPCNPPPSAAPPRIYMNKIFDI 178

Qy 195 -----PLD-----DLK 200
: | : : : : | : : : :
Db 179 KQDERVSMGQNGDLYFANVLTSNHSOYICNAHFPPTRTIIQKEPIDLRVKPTNSMIDRK 238

Qy 201 ---AMSPGASSRVRIVDGNLLIS----- 221
: | | : : | : : : : :
Db 239 PRLFFPTNSSRLVALQGSLLILECIAEGFPTPTIKWLHPSDPMPTDRVIYQNHNTLQL 298

Qy 222 -NVEPIDEGNYKICIAQLVGTRESSYAKLIVQKPYFMKEPKDQVMLIGQTATFHC SVGG 280
| | : | : | : | : : | : : : : : : : : : :
Db 299 LNVGEEDDGEYTCLEANSLSARHAY-YVTVEAAPYWLQKQSHLYGPGETARLDCQVQG 357

Qy 281 DPPPKVLWKKKEGNIPVSRARILHDEK-----SLEISNITPTDEGTVC EAHNNVGQIS 334
| : : : : : : : | : | : | : | : | : | :
Db 358 RPQPEITWRIN---GMSMETVKNQDKYRIEQGSLILSNVQPTDTMTVQCEARNQHGLLL 413

Qy 435 ARASL-IVHAPPNFTKRPSNKKVGLNG-VVQLPCMASGNPPSVFVWTEGVTLMFPNSS 392
| : : : | : : : : | : | : | : | : : : :
Db 414 ANAYIYVQVLPARILTQDNQTYMAVEGSTAYLLCKAFGAPVPSVQWLDEEGTTVL----Q 469

Qy 393 HGRQYVAADGTLQITDVRQEDEGYVCSAFSVVDSSTVVRVFLQVSSVDERPPPIIQIGA 452
: | : | : | : : | : | : : : : : | : : :
Db 470 DERFFPYANGTSLIRDQANDTGRYFCQAANDQNNVTILANLQVKEATQ-----ITQGR 524

Qy 453 NQTLPGKSVATLPCRATGNPS--PRIKWFHDHGAQA---GNRYSIIGSSLRVDDLQLS 507
: | : | : | : : | : | : : : : : | : : :
Db 525 SAIEKKGARVFTTCQAFDPSLQASITWRGDRDLQERGDSDKY-PIDGKLVIQSLDYS 583

Qy 508 DSGTYTCTASGERGET-SWAATLTVEKPGSTSLHRAADPSTYPAPPGTPKVLNVSRSTSIS 566
| | : | : | : | : | : | : | : : : : :
Db 584 DQGNYSVASTELDEVESRAQLLVGSGPGVPVHLESDRHL-----LKQSQVH 631

Qy 567 LRWAKSQEKPGAVGPIIGYTYVEYFSPDL-QTGWIVAHRVGDQTQVITISGLTPGTSYVFLV 625
| : : : : | : : : : | : : : : | : : : :
Db 632 LSWSPAEDHN---SPIEKYDIEFEDKEMAPEKWFSLGKVPNGTSTTLKLSYVHYTFRV 688

Qy 626 RAENTQGISVPSGLSNVKTIEADFDAASANDLSAA-----RTLLTGKSVELIDASAI 678
| | : | : | : | : | : : : : : : : :
Db 689 TAINKYGPGEPSVSESVVTPEA---APEKNPVDVRGEGNETNNMVIWKLPLRWDWNPAP 745

Qy 679 NASAVRLEWMLHVSADKEYVEGLRIHYKSDASVPSAQYHSITVMDASAEVFNGLNKYTK 738
| : | : : | : : : : | : : : : | : : : :
Db 746 QIQ-YRVQWR-----PQKQETWRKQTVSDPFLVSNSTIFVP 782

Qy 739 YEFLTPFFETIEGQPSNKTALTY----EDVPSAPPDNIQIGMYNQTAGWRVTPPSQ 794
| | : : : | : : : | | : : : | : : : |
Db 783 YEIKV---QAVNQKGPEQPQVIGYSGEDYQVSELEDITIFNSSTLVWRVVDLA 838

Qy 795 HNGNLYGYKI-----EVSAGNTMKVLNMTLNATTSVLLNLTGAVYSVRLNS 845
| : | : : : | : : : : : : | : : : : :
Db 839 QVKGHLKGYNVTYWKGSQRKHSKRHIK--SHIVVPANTTSAILSGLRPYSSHYVEVQA 895

Qy 846 FTKAGDGP-----YSKPISLFMDPTH-----HVHPRAHPSGTHDG--- 881
| | | : | : : | : | : : | : | : :
Db 897 FNGRGLGPASEWTFSTPEGV---PGHPEALHLEQSDTSLLLHWQPLSH-NGVLTGYLL 952

Qy 882 -RH--EQG-----DLTYHNNGNIPPGDINPTTHKKTDTYLSGPWLMVLVCIVL 926
| | : : : : | : | : : : : : : : : :
Db 953 SYHPVEGESKDQLFFNLSDPELRTHNLNLNP-DLQYRFQLQATTQGGGPEAIVREGGT 1011

Qy 927 LVLV-----ISA-----AISWYFK-----RKHQMTKELGHSVSDNEITALNIN 967
: | : | : | : | : | : | : | : : : :
Db 1012 MALFGKPDFGNISATAGENYSVSVWVPRKGCNFRFHILFALPEGKVPSPDHQPPQYVS 1071

Qy 968 SKESLWIDHHRGWR-TADTDKDSGLSEKLLSHVNSQSNYNSDGGTDYAEVDT 1021
: : : : | : | : : : : : : : : : :
Db 1072 YNQSSYTQ---WNLQPDTKYBIHLIKEKVLHLDVKTN-----GTGPRVST 1116

Search completed: January 22, 2001, 12:24:13
Job time: 1930 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:26:30 ; Search time 162.41 Seconds
(without alignments)
277.386 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPMPENHAIARSTSTNN.....SCLYAEAGEAPRQMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	759.5	10.2	1377	1	NEO1_RAT	P97603 rattus norv
2	731.5	9.8	1493	1	NEO1_MOUSE	P97798 mus musculus
3	710.5	9.6	1461	1	NEO1_HUMAN	Q92859 homo sapien
4	677	9.1	1443	1	NEO1_CHICK	Q90610 gallus gall
5	661	8.9	2012	1	DSCA_HUMAN	O60469 homo sapien
6	658	8.9	1040	1	AXO1_RAT	P22063 rattus norv
7	644.5	8.7	1040	1	AXO1_HUMAN	Q02246 homo sapien
8	641.5	8.6	1260	1	CAML_MOUSE	P11627 mus musculus
9	639.5	8.6	1447	1	DCC_MOUSE	P70211 mus musculus
10	637.5	8.6	1447	1	DCC_HUMAN	P43146 homo sapien
11	633.5	8.5	1257	1	CAML_HUMAN	P32004 homo sapien
12	627	8.4	1259	1	CAML_RAT	Q05695 rattus norv
13	626.5	8.4	1036	1	AXO1_CHICK	P28685 gallus gall
14	624	8.4	1284	1	NRCA_CHICK	P35331 gallus gall
15	614.5	8.3	1018	1	CONT_HUMAN	Q12860 homo sapien
16	611	8.2	1239	1	NRG_DROME	P20241 drosophila
17	602.5	8.1	1020	1	CONT_MOUSE	P12960 mus musculus
18	595.5	8.0	1010	1	CONT_CHICK	P14781 gallus gall
19	547	7.4	4393	1	PGBM_HUMAN	P98160 homo sapien
20	546	7.4	1897	1	PTPF_HUMAN	P10586 homo sapien
21	543	7.3	1912	1	PTPD_HUMAN	P23468 homo sapien
22	540.5	7.3	1266	1	NGCA_CHICK	Q03696 gallus gall
23	530.5	7.1	2029	1	LAR_DROME	P16621 drosophila
24	514.5	6.9	1091	1	NCA1_CHICK	P13590 gallus gall
25	499	6.7	1115	1	NCA1_MOUSE	P13595 mus musculus
26	497	6.7	3707	1	PGBM_MOUSE	Q05793 mus musculus
27	480.5	6.5	1070	1	PTK7_HUMAN	Q13308 homo sapien
28	460	6.2	1088	1	NCA1_XENLA	P16170 xenopus lae
29	456	6.1	1051	1	PTK7_CHICK	Q91048 gallus gall
30	449.5	6.1	858	1	NCA1_RAT	P13596 rattus norv
31	449	6.0	853	1	NCA1_BOVIN	P31836 bos taurus
32	448	6.0	1092	1	NCA2_XENLA	P36335 xenopus lae
33	439	5.9	848	1	NCA1_HUMAN	P13591 homo sapien

34	437	5.9	837	1	NCM2_MOUSE	Q35136 mus musculus
35	431.5	5.8	761	1	NCA2_HUMAN	P13592 homo sapien
36	431	5.8	725	1	NCA2_MOUSE	P13594 mus musculus
37	415	5.6	837	1	NCM2_HUMAN	Q15394 homo sapien
38	414.5	5.6	1913	1	KMLS_HUMAN	Q15746 homo sapien
39	396.5	5.3	1906	1	KMLS_CHICK	P11799 gallus gall
40	355	4.8	811	1	FS22_DROME	P34083 drosophila
41	353.5	4.8	873	1	FS21_DROME	P34082 drosophila
42	349	4.7	2481	1	UN52_CAEEL	Q06561 caenorhabdi
43	323	4.3	898	1	FAS2_SCHAM	P22648 schistocerc
44	314	4.2	1367	1	VGR2_MOUSE	P35918 mus musculus
45	314	4.2	1694	1	SN_MOUSE	Q62230 mus musculus

ALIGNMENTS

RESULT 1

NEO1_RAT
ID NEO1_RAT STANDARD; PRT; 1377 AA.
AC P97603;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR (FRAGMENT).
GN NEO1 OR NGN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97015074; PubMed=8861902;
RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-Y.,
RA Culotti J.G., Tessier-Lavigne M.;
RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor";
RL Cell 87:175-185(1996).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U68726; AAB41100.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal.
FT NON_TER 1 1
FT SIGNAL <1 2 POTENTIAL.
FT CHAIN 3 1377 NEOGENIN.
FT DOMAIN 3 1074 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1075 1095 POTENTIAL.
FT DOMAIN 1096 1377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 36 105 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 135 197 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 232 296 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 324 386 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 405 502 FIBRONECTIN TYPE-III.

```

ID  NEOI_MOUSE          ,STANDARD;      PRT;   1493 AA.
AC  P97798;
DT  01-OCT-2000 (Rel. 40, Created)
DT  01-OCT-2000 (Rel. 40, Last sequence update)
DT  01-OCT-2000 (Rel. 40, Last annotation update)
DE  NEOGENIN PRECURSOR.
GN  NEOI OR NGN.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN  [1]
RP  SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC  TISSUE-BRAIN;
RX  MEDLINE=97407661; PubMed=9264410;
RA  Keeling S.L., Gad J.M., Cooper H.M.;
RT  "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RT  expressed widely in the adult mouse and during embryogenesis.";
RL  Oncogene 15:691-700(1997).
CC  -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC  TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC  DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC  MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC  -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC  -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
CC  AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC  ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.

```

```
Qy 215 GGNLLISNV-----EPIDEGNYKIA---QNLVGTRRESSYAKVLIVQVKPYFMKEPKDQM 266
   |:| |||| : | ||| |:|:| || | | ||| | |:| :
Db 117 DGSFLISNVVHSHKNKP-DEGFYQC VATVDNL-GTIVSR T AKTLVAGLPRFTSQPEPSSV 174

Qy 267 LYGTQTATFHCSVGGDP PPKVLWKKEBGNIPVSRARILHDEK-----SLEISNITPTDE 319
   |:| :| :| | | | |:| : | :| |:| : | ||| | |
Db 175 YVGSAILNCEVNADLVPFVRWEQ-----NRQPLLLDRIVKLPSGTLVISNATEGDG 227

Qy 320 GTTYCEAHN-NVGOISARASLIVHAPPN-----FTKRPSN--KKVGLNGVVQLPCMASG 370
   || | | : : | | | | | | | | | | | | | | | |
Db 228 GLYRCIVEGGSPKPKFSDEAE LKVLPDPEEIVDLVFLMRPSSMMKVGTQS AV--LPCV VSG 285

Qy 371 NPPPSVFWTKGEVSTLMFPNNSSHGRQYVAADGTLQITDVQRQEDEGYVCASFVSVDSSTV 430
   | | | | | : : | | | : | | |:|:| | :| | | : | :
Db 286 LPAPVVRWMK---NEEVLDTESSGRLVLLAGGCLEISDVTEDDAGTYFC---IADNGNK 338

Qy 431 RVFLQVSSVDERPPPIIQIGPANQTLPGKS VATLPCRATGNPSRIKWFDHGAVQAQNR 490
   | | | | | : | | | | | | | | | | | | | | | |
Db 339 TVEAQAE LTVQVP PGFLK-QPANIYA HESMDIVFECEVTGKTPTTVKWKXGNDVIPSDN 397

Qy 491 YSIIQGSSLRVDDLQLSDSGSYTTCTASGERGETSWAATLTIV-----EKPGST 537
   : |:| : |:| | | | | | | | | | | | | | | : |
Db 398 FRIVKEHNQLVLGVKSDEGFYC I AENDVGNAQAGAQLIILEHDVAIPTLPPTSLTSAT 457

Qy 538 SLHRAADPSTYPAPPGTPKVL NVSR TS---ISLRWAKSQE KPGA VGPIIGITYVEYFSPDL 594
   : | | | |:| | | | : | | | | | | | | | | | :
Db 458 TDHLA--PAT TGPLSPAPRDVASLVSTRFIKLTWRTPASDPH--GDNLTYSVFYTKEGV 513

Qy 595 QTGWIVAHRVGD TQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTI EADFDAAS 654
   : : |:| ||| | | | | | | | | | | | | | | | :
Db 514 DREVENTSQPGEMQVTIQNLMPATVIFYKMAQNKHG-----SGESSAPLRVE----- 562

Qy 655 ANDLSAARTLLTGKSVEL-----IDASAINASAVRLEWLHVSADEKYVEGLRIHYKDA 708
   | :| | | | | | | | | | | | | | | | | | | |
Db 563 -----TOPEVQLPGPAPNIRAYATSPTSITVTWETPLSGNGE-IQ NYKLYMEK 610

Qy 709 SVPSAQYHSITVMDASAESFVGNLK KYTKYEFFLTPFFETIEGQPSNSKTALTYEDVPS 768
   | :| :| :| :| :| | | | | | | | | | | | | | |
Db 611 GTDKEQ----DIDVSSH SYTINGLKKYTEYSPRVVAYNKHGPGVSTQDVAVRTLSDVPS 665

Qy 769 APPDNIQIGMNQTAGVWRWTPPPSOHHNGNLYGKIEVSAGNTMKVLAMTNLATTSV 828
   | | | :| :| :| | | | | | | | | | | | | | | |
Db 666 AAPQNLSEVRNKSIVIHWQPSSSTQNGOITGYKIRYKASKRS DVTIELVTGTQLSQ 725

Qy 829 LLNLLTTGAVISVRLNSFTKAGDGPYSKPIS--LFMDPTHVHVHPRAHPSGTHDRHEGQ 886
   |:| :| | | | | | | | | | | | | | | | | |
Db 726 LIEGLDRGTENFRVAALT VNNGTGPATDWLSAETFESDLDET R VEV-PSSLH----- 777

Qy 887 DLTYHNNGNIPPGDINPTTHKKT DYLSGPLMWLVLCIVLLVLVISA AISMVYFKRKHKM 946
   : | | | | | | | | | | | | | | | | | | | |
Db 778 -----VRP-----LVT SIVVSTPPPENQNI V 798

Qy 947 TK--ELGHL SVVSDNEITALNINSKESLWDIHHRGWRTADTDKDSGLSES KLLSHVNSSQ 1004
   : |:| :| :| :| | | | | | | | | | | | | | :
Db 799 VRGYAIGY-----GIGSPA HQIKVDYKQRYIYT ENL DPS-----SHYVITL 840

Qy 1005 SNYNNSDGG-----TDYAEVDTRNLTFYCNCRKSPDNPTPYATTMI--IGTSS 1050
   |:| | | | | | | | | | | | | | | | | | | |
Db 841 KAFNNVGEGIPLYESA VRPHDTSEVDLFVI-----NAPYTPVPDPTPMMPVGVQA 893

Qy 1051 SETCTKTT SIS-ADKDSGTHSPSYDA-----FAGQVPAPV PVVKS-----NYLOYPVPEP 1097
   | | | | | | | | | | | | | | | | | | | |
Db 894 SILSHDTIRITWADNSLPKHQKITDSRYTYVRKTNIPANTKYKANATTL SYLVTGLKP 953

Qy 1098 INWSEF-----LPPPPPEHPPPSSTYG AQGSPES-----SR 1128
   | | | | | | | | | | | | | | | | | | | |
Db 954 NTLYEF SVMVTKGRSS TWSMTAHGAT FELVPTSPPKDVTVVSK EGKPTIIVNWQPPE 1013

Qy 1129 KSSKSAGSGI--STNQSI LNASHSSSSGGFSAWGVSP-----QYAVA 1169
   : | | | | | | | | | | | | | | | | | | | |
Db 1014 ANGKITGYLIYYSTD--VNAEIHD-----WVIEPVVGNRLTHOI QELTLDTPYYFK 1063
```

```

Qy      124 EQDGGEYWCVAKNR-----VGQAVSRHASLQIAVLRDD----- 156
      |::| |::| |::| |::| |::|
Db       4 EREAGRLCLTSSSRRCPPPLLLLLLPLLLLRPASGAATKSGPRRQSGCASVRFFTP 63

Qy      157 --RVPEKDTFVAKGETALLECQPKPIEPTLIWKDGVPLDDLKAMSFAGSSRVRIVD 214
      ||||| |::| |::| |::| |::| |::|
Db      64 FYFLVEVPDTLSVRGSSVILNC-AYPESPINTEIWKDKGT-----FNLESDDRRRLQIP 116

```

Qy 1170 CPPENVISNPLSAVAGGTQNRQITPTNQHPQLP---AYFATTGPGGAVP----- 1217
 | | : :: | | || ::| : | | : |
 Dh 1063 IOARN---SKGMCNPSFAVOER---TPKADSSKWPNDQALGSAGKSGSRPLDLSGYKPPM 1117

Qy 1218 -----PNHLP-----FATQRHAASEYQAGLNAARCA 1243
|: | | |: | : :
Db 1118 SGSNSPHGSPTSPDLSNMLLVIIIVSVGVITIVVVVIAVEFCRRTTSHOKK----- 1168

Qy 1244 QSRACNS-----CDALATPS-----PMQP---PPVPVPWEGWYQVHPNSH 1281
: ||| : | :|| ||| : ||
Db 1169 KRAACKSVNGSHKYKGNCKDVKPDDLWIIHHERLELKPIDKSPDPNPMVTD--TPIPRNSO 1226

Qy 1282 PMHPTSSNHQIYQCSSECDHSRSSQSHKRLQLLEEHS--SAQRGGHRRRAPVVPQ 1338
: | : | | | : | | | | : |
Db 1227 DITPV-----DNSMSDSNIHRRNSYRGHESEDSMSLTAGRRGMRPKMMP 1271

RESULT 3

```

NEO1_HUMAN
ID   NEO1_HUMAN          STANDARD;   PRT;   1461 AA.
AC   Q92859; 000340;
    01-OCT-2000 (Rel. 40, Created)
    01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NEOGENIN PRECURSOR.
GN   NEO1 OR NGN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE=97236653; PubMed=9121761;
RA   Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT   "Identification and characterization of neogenin, a DCC-related
RT   gene.";
RL   Oncogene 14:1129-1136(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE=97312699; PubMed=9169140;
RA   Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA   Korenberg J.R., Dreyer W.J.;
RT   "Molecular characterization of human neogenin, a DCC-related protein,
RT   and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT   q23.";
RL   Genomics 41:414-421(1997).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC   TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC   DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC   MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC   LINES.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC   TUMOR SUPPRESSOR PROTEIN DCC.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collabor
CC   between the Swiss Institute of Bioinformatics and the EMBL outstat
CC   the European Bioinformatics Institute. There are no restrictions on
CC   use by non-profit institutions as long as its content is in no
CC   modified and this statement is not removed. Usage by and for comme
CC   entities requires a license agreement (See http://www.isb-sib.ch/anno
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U61262; AAB17263.1; -.
DR   EMBL; U72391; AAC51287.1; -.
DR   MIM; 601907; -.
DR   HSSP; P02751; ITTG.

```

```

DR INTERPRO: IPRO01777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; ig; 4.
DR PRINTS: PR00014; FNTYPEIII.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1461 NEOGENIN.
FT DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1106 1126 POTENTIAL.
FT DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 67 136 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 166 228 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 263 327 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 355 417 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 436 533 FIBRONECTIN TYPE-III.
FT DOMAIN 536 629 FIBRONECTIN TYPE-III.
FT DOMAIN 630 729 FIBRONECTIN TYPE-III.
FT DOMAIN 735 829 FIBRONECTIN TYPE-III.
FT DOMAIN 850 950 FIBRONECTIN TYPE-III.
FT DOMAIN 951 1052 FIBRONECTIN TYPE-III.
FT DOMAIN 1118 1121 POLY-VAL.
FT DISULFID 74 129 BY SIMILARITY.
FT DISULFID 173 221 BY SIMILARITY.
FT DISULFID 270 320 BY SIMILARITY.
FT DISULFID 362 410 BY SIMILARITY.
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 1248 1300 MISSING (IN ISOFORM 2).
FT CONFLICT 168 168 G -> N (IN REF. 2).
SQ SEQUENCE 1461 AA; 159958 MW; 7AAE897E569635A21 CRC64;

```

Query Match 9.6%; Score 710.5; DB 1; Length 1461;
Best Local Similarity 22.3%; Pred. No. 3.1e-28;
Matches 313; Conservative 184; Mismatches 564; Indels 341; Gaps 48;

Qy 157 FRVEPKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPPLDLKAMSGASSRVRIVDGG 216
| | | | | : : : | : | : | : : : : :
pk 55 RIUNYDI SUDGSHI NGS - AYSGEPIEITKKCT-----RIUYSDDPRL DPG 107

Qy 217 NLLISNV-----EPIDEGNYKCIAQ-NLVGTRESSYAKLIVQVKPYFMKEPKDQVMLYG 269
: ||||| :| ||| |:|:| :|| | |||| | | :|: : |
Db 108 SLFSTNNVVSHSKNKP-DEGVYOCVATVESIGTISRPAKLIVAGIPRETSEPPSSSYAG 166

Qy 270 QTATFHCSVGGDPPPKVLWKKEGNIPVSRARILHDEK-----SLEISNITPTDEGT 322
| : | | | | : : : : | : | : : | | | | |
Db 167 NGAILNCEVNADLVPEVRWEO-----NROPLLLDDRVIKLPSGMLVISNATEGDGGLY 219

Qy 323 VCEAHN-NVGQISARASLIVHAPPN-----FTTKPSNKKVGLNGVQLPCMASGNPPPS 375
| : : | | | | | | : | : | : | : | : | :
Db 220 RCVVESGGPKYSDEVELKVLDPDEVISDLVFLKSPSLVRVIGODVVLPCVASGLPTPT 279

Qy 376 VFVTK--EGVSTLMPNSSHGRQYVAADGTLQITDVRQEDEGYVVCSAFSVVDSSTVRVF 433
:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| :
Db 280 IKWKNNEEALDT----ESSERLVLAGGSLEISDVTEDDAGTYFC---IADNGNETIE 330

Qy 434 LQVSSVDERPPPIIQIGPANQTLPGKGSVATLPCRATGNPSPRIKWFHDGHAVQAGNRYSI 493
| : | :: | | | | | | : | : | : | : | : |
Db 331 AQAEALTVOAQPEFLK-OPTNIYAHESMDIVFECEVTGKPTPTVKVVKNGDMVPSDYFKI 389

Qy 494 IQGSSRLVRDDLQLSDSGTYTCTASGERGETSWAATLTVEKPGSTSLHRAADPTYPAPP 553
:: :|: | || | | : | : : | | (: : |
Db 390 VKEHNLOVLGLVKSDEGFYOCIAENDVGNAQAGAQLIILE-----HAPATTGPLSAPR 443

Ov 554 TPKVLNVSRTSISLRWAKSOEKP GAVGPIIGYTVEYFSPDLQGWIVA AHRVGD TQVTIS 613

Db 444 DVVASLVSTRFKLWRTPASDPH--GDNLTYSVFYTKEGIARERVENTSHPGEMQVTIQ 501
Qy 614 GLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFAASANDLSAARTLLTGKSVEL- 672
Db 502 NLMPATVYIFRMAQNKHG---SGESSAPLVE-----TQPEVQLP 539
Qy 673 -----IDASAINASAVRLEWMLHVSADKEYEGLRIHYKASVPSAQHSITVMDASAES 727
Db 540 GPAPNLRAYASPTSITVTWETPVSGNGE-IQNYKLYMEKGTDEKQ-----DVDVSSH 593
Qy 728 FVVGNLKKYTKYEFFLPFFETIEGQPSNKTALTIEDVPSAPPDNIQIMYNTAGWVR 787
Db 594 YTINGLKRYTEYSFRVAYNKGPGVSPDVAVRTLSQVSAAPQNLSEVNRKSKSIMH 653
Qy 788 WTPPPSQHNGNLYGYKIEVSAGNTMKVLNMTLNTATTSVLLNLTGAVISVRLNSFT 847
Db 654 WQPAPATQNGQITGYKIRYKASRKSVDVETLVSGTQLSLIEGLDRGTENFRVAALT 713
Qy 848 KAGDGPYSKPIS--LFMDPTHVHPRAHPSGTHDRHGQDLTYHNGNIPPQDINPT 905
Db 714 INGTGPATDWSAETFSDLDETRVPEV-PSSLH-----VRP-- 749
Qy 906 HKKTTDYLSGFWMLVLCIVLLVLVISAISMVYFKRKQMTK--ELGLHSVSDNEITA 963
Db 750 -----LVTISVSWTPPENQNIIVRGYAIQY-----G 776
Qy 964 LNINSKESLWIDHHRGWRTADTKDSEKSLSHVNSSQSNYNSDGG----- 1013
Db 777 IGSHPAQTIKVDYKQRYTITENLDP-----SHYVITLKAFFNVGEGIPLYESAATR 828
Qy 1014 --TDYAEVDTRNLTFYCNCRSPDNPTPYATTMI--IGTSSSETCTKTTTIS-ADKDSGT 1068
Db 829 PHTDTSEVDFVI-----NAPYTPVPDPTMPMPVGVQASILSHDIIRITWADNSLPK 881
Qy 1069 HSPYSDA-----FAGQVPAVPVVK-----NYLQYVPEPINWSEF----- 1103
Db 882 QOKITDSRYTYRWKNTIPANTKYKANATTSYLVTLGLKNTLYEFSVMVTKGRSSSTW 941
Qy 1104 -----LPPPEHPPPSPSTGYAQGPSPE-----SRKSSKSAGSGI--STNQSL 1145
Db 942 SMTAHGTFELVPTSPPKDVTVVSKEGKPTIIVNQWQPSEANGKITGYIIYITD--V 998
Qy 1146 NASIH-----SSSSGGFSAMGVSPQYAV-ACPPEN 1174
Db 999 NAEIHDWVIEPVVGNRLTHQIQELTLDTPYFKIQAQNSKGMGPMSEAVQFRTPKADSSD 1058
Qy 1175 VYSNPLSAVAGGTQNRQITPTNQHPQLPAYFATTPGPGGAVPPNHL----- 1221
Db 1059 KMPNDQASGSGKGRSLPDLGSDYKPPMSGSSPHGSPTSPLDNMLLVIIIVSGVITIV 1118
Qy 1222 -----PFATQRHAASEYQAGLNAARCAQSRACNSCDA----LATPSMQPP----- 1263
Db 1119 VVVIIVACTRTTSHQKK-----KRAACKSVNGSHYKGNKDKVPPDLWIHHER 1169
Qy 1264 -----PPVPVPEGWYQPVHPNHPMTSSNHQIYQCSSECDHSRSQSHKRLQL 1315
Db 1170 LELKPIDKSPDPNPIMTDTPIPRNSQDITPV-----DNSMDSNIHQRRNSY 1215
Qy 1316 EEHGSS-----AKQGGHHRRAPIV---VQPCMESENNMLAEYEQRYTSDCCNSSR 1365
Db 1216 RGHESEDSMTLAGRRGMRPKMMFPDQPPQVISAHPIHSLDNPHHFFHSSSLASPAR 1275
Qy 1366 EGDTCSCSEGSCLYAEAGEPAP 1387
Db 1276 -----SHLY-HPGSPWP 1286

RESULT 4
NEOL_CHICK
ID NEOL_CHICK STANDARD; PRT; 1443 AA.
AC Q90610;
DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE=95105243; PubMed=7806578;
RA Vielmetter J., Roman J.M., Dreyer W.J.;
RT "Neogenin, an avian cell surface protein expressed during terminal
RT neuronal differentiation, is closely related to the human tumor
RT suppressor molecule deleted in colorectal cancer.";
RL J. Cell Biol. 127:2009-2020(1994).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
CC AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07644; AAC59662.1; -.
DR HSSP; P80362; 1WTL.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN <1 1090 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1091 1111 POTENTIAL.
FT DOMAIN 1112 1443 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 102 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 132 194 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 229 293 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 321 383 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 422 519 FIBRONECTIN TYPE-III.
FT DOMAIN 522 615 FIBRONECTIN TYPE-III.
FT DOMAIN 616 714 FIBRONECTIN TYPE-III.
FT DOMAIN 720 814 FIBRONECTIN TYPE-III.
FT DOMAIN 835 935 FIBRONECTIN TYPE-III.
FT DOMAIN 936 1037 FIBRONECTIN TYPE-III.
FT DISULFID 40 95 BY SIMILARITY.
FT DISULFID 139 187 BY SIMILARITY.
FT DISULFID 236 286 BY SIMILARITY.
FT DISULFID 328 376 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1443 AA; 158050 MW; 558C6795579C0E26 CRC64;

Query Match

9.1%; Score 677; DB 1; Length 1443;

Best Local Similarity 23.1%; Pred. No. 1.5e-26;
Matches 304; Conservative 171; Mismatches 541; Indels 302; Gaps 49;

Qy 157 FRVEPKDTRVAKGETALTECG-----PPKGIPEPTLIWIKDGPLDDLKAMSGASSRVR 211
| | | | : : : | | : : | : : : :
Db 21 FLVEPMILSVRGASVIMNCSSYCEPPK-----IEWKKGDT-----LLNLVSDRRQ 68

Qy 212 IVDGGLNISNV-----EPIDEGNYKICIAQ-NLVGTRESSYAKLVQKPYFMKEPKDQ 264
: : | : | : : | : | : | : : | : | : | : :
Db 69 LLPDGSLLINSVVHSHKNKP-DEGYTQCVAIVESLGSIVSRITAKLTAGLPRTISQPELS 127

Qy 265 VMLYGQTATFHCSVGGDPPPKVLWKKKEGNIPIVSRARILHDEKSLEISNITPTDEGYVC 324
: : | : | : | : : : : : : : : : : : : :
Db 128 SVYKGSAILNCEVNDLAPFVRWEQDRQLSLDDRVFKPLSGALLIGNATDTDGGFYRC 187

Qy 325 EAHN-NVGQISARASLIVHAPPN-----FTKRPSN--KKVGLNGVQLPCMASGNPPPS 375
: : | : | : : : : : : : : : : : : :
Db 188 VIESGGTPKYSEEALKILPDEEPQSLVFRVQPSLTKVTGQNAV--FPCVAGGFTPY 245

Qy 376 VFWKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQDEGYVCSAFVSDSSTVRVFLQ 435
| | | | : : : | : | : | : | : | : : : :
Db 246 VRWTKNGEELI---TEDSERFALRAGSLLISDVTEEDVGTTC---IADNENETIEAQ 298

Qy 436 VSSVDERPPPIIIGPANQTLPGKSVATLPCRTGNPSRIKWFHDHAGVAGNRYSIQ 495
: : | : : | : : | : | : | : | : | : : : :
Db 299 AELAVQVPEFLK-RPANIYAHESMDIVFECEVTGKPTPTVKWVKNQDVVPSYFKIVK 357

Qy 496 GSSLRVDDLQSDSGTCTTASGERGETSWAATL-----TVEKPGSTSLHRAAD--- 544
: : | : | : | : : | : : | : : | : : :
Db 358 EHNQLVGLVKSDEGFYQCIENQVNAQAQALILDLDAIPTLPPTSLTSATNDHLA 417

Qy 545 PSTYAPPGTPK---VLNVSRTSISLRWAKSQEKGAVGPIIGTYVEYFSPDLQTGWIVA 601
: : | : : | : | : | : : | : : : : : : :
Db 418 PATGTPLTAPRDVVATLVSTRFIRLTWRTPVSDP--QGDNLTYISIFTYKGINRERVEN 475

Qy 602 AHRVGTQVTISGLTPTGTSYVFLVRAENTQGISVPSGLSNVIRTEADFAASANDLSAA 661
| : | : | : | : | : | : | : | : | : | :
Db 476 TSRPGETQVMIQNLMPETVYVFRVVAQNKHGHESSAPLKVATQPEVQLPGAPN----- 530

Qy 662 RTLLTGKSVELIDASAINASAVRLEMLHVSADKYEGLRIHYKASVPSAQYHSITVM 721
| : | : : : | : : : : : : : : : : : :
Db 531 -----IRAYAGSPSTVTVTWETPLSGNGB-IQNYKLYMEKGQDSEQ-----DV 573

Qy 722 DASAEFVGNLKKYTKYEFLLTPFFETIEGQPSNKTALTYEDVPSAPPDNIQWIMYQ 781
| : : : | : | : | : : : : : | : | : | : :
Db 574 DVAGLSYITGLKKYTEYSFRVWAYNKHGPCVSTQDVVVRLLSDVPSAAPQLNLEARN 633

Qy 782 TAGVWRVTPPPSQHNGNLYGYKI---EVSAGNTMKVLANMTLWATTSVLLNLTGAV 838
: : | : | : | : | : | : | : | : | : | :
Db 634 KSIMLHWQPPAGTHSGQITGYKIRKRVK---RKSDVTSVGGTQLFQLEGLERGTE 689

Qy 839 YSVRLNSFTKAGDGPYKSPIS--LFMDPTHVHPRAHPSGTHDRHEGDLTYHNGNI 896
: : | : : | : | : : | : | : | : | : :
Db 690 YNFRIAAMTVNGTGPATDWSAETPESDLDESVPYV-PSSLH----- 731

Qy 897 PPGDINPTTHKKTIDYLSGFWLWLVLCIVLLVISAISMVYFKRKHQMTK--ELGHS 954
: : | : | : | : | : | : | : | : : : :
Db 732 ----VRP-----LVTISVSVTPPENQNIIVRGYAGY-- 760

Qy 955 VVSDNEITALNINSKESLWIDHHRGWRTADTKDSGLSEKLLSHVNSSQSYNNNSDGG- 1013
: : : : : : : : : : : : : : : : : :
Db 761 -----GIGSPAQTIKVDYKQRYTITENLDPS-----SHYVITLKAFNNVGEI 804

Qy 1014 -----TDYAEVDTRNLTFYNCRKSPDNPTPYATTMI--IGTSSSETCTKTTSI 1060
: : | : | : : | : | : | : | : | : : :
Db 805 PLYESAVTRPHSDTSEVDLFI-----NAPYTPVPDPSPMPPVGVQASILSHDTIRI 857

Qy 1061 S-ADKDSGTHSPYSDA-----FAGQVPAVPVKS-----NYLQYPVEPINWSEF--- 1103
: : | : : | : : : : | : | : : : : :
Db 858 TWADNSLPKNQKITDARYTYRWKTNIPANTKYKTANATLSLVTGLKPNLTLYEFVMV 917

Qy 1104 -----LPPPEHPPPSSYGYAQGPSPE-----SRKSSKSAGSGI 1138
| : | : | : | : | : | : | : | : |

Db 918 TGRSSSTWSMTAHGTTTFELVPTSPKDVTVVSKGKPRTIIVNQWPPSEANGKITGYII 977

Qy 1139 --STNQSLNASHSSSSGGFSAMGWSPQYAVACPPENVYSNPLSAVAGGTQNRQYITPT 1196
| : : | : | : | : | : | : | : : :
Db 978 YSTD--VNAEHD-----WVIEP-----VVGNRIT-----HQIQEL 1007

Qy 1197 NQHPPQLPAYFATTGPGGAVPPNHLFPATQRHAASEYQAGLNAARCAQSRACNSDALAT 1256
| : | : | : | : | : | : | : | : :
Db 1008 TLDTPYYFKIQARNKGMGPSEAVQFRTPKAESD-----KMPNDQASGSAGKGR 1059

Qy 1257 PSPMQP--PPFPVPEGWYQPVHPNSHPMPTSSNHQIYQCSSE-----CS 1300
| : | : | : | : | : | : | : | : :
Db 1060 PVDVGPDKYKPLSGSNS-----PHGSPTSLDLSNMLLVIIVSVGVITIVIVIVAVFT 1113

Qy 1301 DHRSSQSHKRLQLEEHGSSAKQKRG-----HHR-----RAPVVPQCM 1340
: : | : | : | : | : | : | : : : :
Db 1114 RRTTSHQKKRAACKSVNGSH-KYKGNKDVKPPDLWIHHERLELKPDKSPDPNPM 1170

RESULT 5

DSCA_HUMAN

ID DSCA_HUMAN ; STANDARD; PRT; 2012 AA.
AC O60469; O60468;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE PRECURSOR (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system."
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
RN [3]
RP FUNCTION.
RA Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion."
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
CC -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

FT	CARBOHYD	28	28	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1562	1571	NFATL1YDGS -> KEAARCKEFS (IN SHORT

Db 1176 RSEOI--FTRTKEDVPGP---PAG----- 1194

```

RESULT 16
AXO1_RAT
ID AXO1_RAT STANDARD; PRT; 1040 AA.
AC P22063;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1).
GN TAX1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC TISSUE=SPINAL CORD;
RX MEDLINE=90199890; PubMed=2817872;
RA Furley A.J., Morton S.B., Manalo D., Karagogeos D., Dodd J.,
RA Jessell T.M.;
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily
RT member with neurite outgrowth-promoting activity.";
RL Cell 61:157-170(1990).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS, AND IN ADULT
CC BRAIN, SPINAL CORD AND CEREBELLUM.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY EXPRESSED ON A SUBSET OF AXONS
CC IN THE DEVELOPING RAT NERVOUS SYSTEM.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31725; AAA42201.1; -.
DR PIR; A34695; A34695.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.

```

DR	PFAM; PF000041; fn3; 4.	
DR	PFAM; PF000047; ig; 6.	
KW	Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;	
KW	Cell adhesion; Repeat.	
FT	SIGNAL	1; 30
FT	CHAIN	31; ?1015
FT	PROPEP	?1016; 1040
FT	DOMAIN	56; 120
FT	DOMAIN	150; 218
FT	DOMAIN	256; 315
FT	DOMAIN	343; 404
FT	DOMAIN	435; 497
FT	DOMAIN	525; 596
FT	DOMAIN	608; 614
FT	DOMAIN	613; 708
FT	DOMAIN	716; 811
FT	DOMAIN	818; 910
FT	DOMAIN	911; 1005
FT	SITE	796; 798
FT	CARBOHYD	78; 78
FT	CARBOHYD	200; 200
FT	CARBOHYD	206; 206
FT	CARBOHYD	463; 463
FT	CARBOHYD	479; 479
FT	CARBOHYD	500; 500
FT	CARBOHYD	527; 527
FT	CARBOHYD	777; 777
FT	CARBOHYD	832; 832
FT	CARBOHYD	920; 920
FT	CARBOHYD	942; 942
SQ	SEQUENCE	1040 AA; 113042 MW; 6E707EF6614CB4FB CRC64;

Query Match 8.9%; Score 658; DB 1; Length 1040;
Best Local Similarity 24.1%; Pred. No. 8.6e-26;
Matches 251; Conservative 145; Mismatches 433; Indels 214; Gaps 36;

```

Qy 22 SRSRSSRMWLLPAWLLLVLA---SNGLPVARGQYQSPRIEHTPLDVKK--NEPAT 74
   :|::||::||::||::||::||::||::||::||::||::||::||::||
Db 5 ARKKASLLLVLATVALVSSPGWSFAQGTPTFG---PIFEEQPIGLLFPEESAEDQVT 60

Qy 75 LNCXVEGKPEPTIEWFKDGEVSTNEKSHRVQKDGALFFYRTMGKKEQDGGEYWCVA 134
   ||::||::||::||::||::||::||::||::||::||::||::||::||
Db 61 LACRARASPPATYRWKMNG--TDMNLEPGSRHQLMGGNLVI---MSPTKTQDAGVYQCLA 115

Qy 135 KNRVQAVSRHASLQIAVLRDDFVPEKDRVAKGETALLECGPPKGIPEPTLIWI---- 190
   ||||::||::||::||::||::||::||::||::||::||::||::||::||
Db 116 SNPVGTVYSKIAVLRFGLQEFSEKREDPVKTHEGWGMLPCNPAPHYGLSYRWLLNEF 175

Qy 191 KDGVPDDLKAMSGASSVRIVDGGNLLISNVPEIDEGNYKCIAQNLV--GTRE--SSY 246
   :||::||::||::||::||::||::||::||::||::||::||::||::||
Db 176 PNFIPDTGRHFVS-----QTTGNLYIARTNASDLGNYSCLATSHMDFSTKSVSFKF 226

Qy 247 AKLIV-----QVKPYFMKEPKDQVMLYGQTATFHCSSVGDDPPPKVLWKEEGNIP 296
   ||::||::||::||::||::||::||::||::||::||::||::||::||
Db 227 AQLNLAAEDPPLFAPSIAKRF---PPETYALVGQVITLFCFAGNPVPVRIKRWVDGSL 283

Qy 297 VSRARILHDEKSLEISNITPTDEGTYVCEAHNVVGQISARASLIVHAPPNFKRPSNKKV 356
   |::||::||::||::||::||::||::||::||::||::||::||::||
Db 284 POWATA---EPTLQIPSVSFEDEGTYCEAENSKGRDVTQGRIIVQAQPEWLKVISDETA 340

Qy 357 GLNGVQQLPCNASGNPPPSVFWTKEGVSTLMPNSSHGROVVAADGTLQITDVRQEDEGY 416
   :||::||::||::||::||::||::||::||::||::||::||::||::||
Db 341 DIGSNLRWCCAAGKPRPMVRWLRNGE-----PLASQNRVEVLA-GDLRFSKLSLEDSGM 394

Qy 417 YVCSAFSVSDS--STVRVFLQVSSVDERPPPIIQTIGPANQTLPKGVSATLPCRATGNPSP 474
   ||||::||::||::||::||::||::||::||::||::||::||::||::||
Db 395 YQCVAKNKHGTIYASAEALVAQALAPDFRQNPVRRLLIPA----ARGEISILCQRAAPKA 450

Qy 475 RIKWFHDGHAVQAGNRYSIQGGSSLRVDLQLSDSGTYTCTASGERGETSWAATLTVE-- 532
   ||::||::||::||::||::||::||::||::||::||::||::||::||
Db 451 TILWSKGETILGNSTRVTVTSDGTLIIRNISRSDGKYTCFAENFMKANSTGILSVRDA 510

```

Qy 533 ----KPGSTSL-----HRAADPS---TY----- 548
Db 511 TKITLAPSSADINVGDNLTLCCHASHDPTMDLTFTWLDPPIDFKPGGHYRRASAKET 570
Qy 549 -----PAPPPTPKVLNVSRTSISLRW 569
Db 571 IGDLTILNAHVHRGGKYTCMAQTVVDGTSKEATVLVRGPPGPGVVRDIDGTTVQLSW 630
Qy 570 AKSQEKPGAVGPIIGYTVEYFSPDLQGTWIVAHRV---GDTQVT-ISGLTPGTSYVFLV 625
Db 631 SRGFDNH---SPIAKYTLQARTPPSGKWKQVTRNPNVINEGAETAQVGLGMPMDYEFVR 687
Qy 626 RAENTQGISVPSGLSNVKTIEADFAASANDLSAARTLLTGKSVELIDASAINASAVRL 685
Db 688 SASNILGTGEPSPSSKIRTKEA-VPSVAPSGLSGG---GGAPGELI-----I 731
Qy 686 EWMHLVSADKEYVEGLRIHYKDSVPSAQYHSITVMDASAEFVVGW--LKKYTKYEFL 743
Db 732 NW-TPVSREYQNGDGFYLLSFRQQSSSQRTARVPGADAQYFVYGNDSIQYPTPEFVKI 790
Qy 744 TPFETIEGQPSNSKTALTY--EDVPSAPPDNIQIMYNQTAGVWRTPPPSQHNGNLY 801
Db 791 RSY--NRRGDGPESLTALVYSAEEPRVAPKVVAKSSSEMNVSW-EPVLQDMNGILL 847
Qy 802 GYKIEV-SAGNTMKVLNMTLNTATTSVLLNLTGAVSVRLNSFTKAGDGPYSKISL 860
Db 848 GYEIRYWKAGDNEAAADRVRTAGLTSARVTLGNPKYHVTYRINRAGTGPASPSADA 907
Qy 861 FMDPTHVHVPPRAHPSGT-----HGRHEGQDLTYHNGNIPP 899
Db 908 MT-----VKPPRRPPGNISWTFSSSSLSLKWDPVPLRNESTVTGYKMLYQN----- 955
Qy 900 DINPTTHKKTIDYLSGPWMLV 922
Db 956 DLHPT---PTLHLTSKNWIEIPV 975

RESULT 7

AXO1_HUMAN
ID AXO1_HUMAN STANDARD; PRT; 1040 AA.
AC Q02246;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL GLYCOPROTEIN 1).
RE TAXI1 OR TAG1.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=93145965; PubMed=8425542;
RA Hasler T.H., Rader C., Stoeckli E.T., Zuellig R.A., Sonderegger P.;
RT "cDNA cloning, structural features, and eucaryotic expression of human TAG-1/axonin-1";
RL Eur. J. Biochem. 211:329-339(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=94140354; PubMed=8307567;
RA Tsiotra C.P., Karagogeos D., Theodorakis K., Michaelidis M.T.,
RA Modi W.S., Furlay J.A., Jessel M.T., Papamatheakis J.;
RT "Isolation of the cDNA and chromosomal localization of the gene (TAXI1) encoding the human axonal glycoprotein TAG-1";
RL Genomics 18:562-567(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68274; CAA48335.1; -.
DR EMBL; X67734; CAA47963.1; -.
DR PIR; S28830; S28830.
DR MIM; 190197; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; lg; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1012 AXONIN-1.
FT PROPEP 1013 1040 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 54 118 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 148 216 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 313 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 341 402 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 433 495 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 523 594 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 606 612 GLY/PRO-RICH.
FT DOMAIN 611 706 FIBRONECTIN TYPE-III.
FT DOMAIN 714 809 FIBRONECTIN TYPE-III.
FT DOMAIN 816 908 FIBRONECTIN TYPE-III.
FT DOMAIN 917 1003 FIBRONECTIN TYPE-III.
FT SITE 794 796 CELL ATTACHMENT SITE (BY SIMILARITY).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 1012 1012 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1040 AA; 113393 MW; 254E78DD3C28EFB6 CRC64;

Query Match 8.7%; Score 644.5; DB 1; Length 1040;
Best Local Similarity 23.7%; Pred. No. 4.1e-25;
Matches 245; Conservative 149; Mismatches 416; Indels 223; Gaps 39;

Qy 36 LLLV---LVASNGLPAVRGQYQS--PRIEHPTDLVVK---NEPATLNCKVEGKPEPT 86
Db 11 LLLVAVALVSSAWSSALSGSTTFGPFVEDQPLSVLPFEESTEEQVLLACRARASPPAT 70
Qy 87 IEWFKDGEPEVSTNEKSHRVQFKDGLFFYRTMQGKKEQDGEYWCVAKNRVQAVSRHA 146
Db 71 YRWKMGTEMKLEPGSRH---QLVGGNLVI---MNPTKAQDAGVYQCLASNPVGTVVSREA 125
Qy 147 SLQIAVLRRDPRVEPKDTRVAKGETALLECGPKGIPEPTLIWI---KDGVLDDLKAM 202
Db 126 ILRFGFLQEFSGKEERDPVKAHEGWGVMPCNPAPHYGLSYRWLLNEFPNFIPTDGRHFV 185
Qy 203 SFGASSRVRIVDGNNLSINVEPIDEGNYKICIAQNLV--GTRE--SSYAKLIV----- 251
Db 186: S-----QTTGNLYIARTNASDLGNYSCLATSHMDPSTKSVFSKFAQLNLAAEDTRL 236
Qy 252 ---QVKPYPMKEPKDQVMLYGTATFHCSVGGDPPPKVLWKKKEGNIPIVSRARILHDEKS 308
Db 237 FAPSIKARF---PAETVALVGQVLTLECFAGFNPNVPIKRWKVDGSLSPQWTTA---EPT 290
Qy 309 LEISNITPTDEGTYVCEAHNNVVGQISARASLIYHAPPNFTKRPSNKKVGLNGVQLPCMA 368

Db 291 LQIPSVFDEGTYECAENSGRDTVQGRIVQAQPEWLKVISDTEADIGSNLRWCAA 350
 Qy 369 SGNPPSVFVTKVSTLMFPNSSHGRQYVAADGTLQITDVRQDEGYVCSAFSVVDSS 428
 Db 351 AGKPRPTVRLWRNGE-----PLASQNRVEVLA-GDLRFSKLSLEDGMYQC-----VAENK 400
 Qy 429 TVRVFLQVSSVDERPPPIIQIGPANQTLTLP--KGSVATLPCRATGNPSRIKWFHDHGA 486
 Db 401 HGTIYASAEALAVQALAPDFRLNPFVRLIPAARGGEILIPQCPRAAPKAVVLSKGTETLV 460
 Qy 487 AGNRYSIQGSLLRVDLQSDSGTYTCTASGERGETSWAATLT----- 531
 Db 461 NSSRVTVTPDGTLLIRNISRSDEGKYTCFAENFMGKANSTGILSRDATKITLAPSSADI 520
 Qy 532 -----EKP----- 535
 Db 521 NLGDNLTQCHASHDPTMDLTFTWLTDDPFDKPGHYRRNTVKETIGDLTILNAQLR 580
 Qy 536 -----STSLHRAADPSTY-----PAPPGTPKVLNVSRTSISLRWAKSQEKPAGVGP 581
 Db 581 HGGKYTCMAQTVVDSASKEATVLRVPGPPGPGVVVRDIDGTTIQLSWSRGFDNH---SP 637
 Qy 582 IIGYTVVEYFSP-----DLQGWIAAHRVGTQVIT--ISGLTPGTSTYVFLVRAENTQGIS 634
 Db 638 IAKYTLQARTPPAGKWKQVRTN---PANIEGNAETAQVLGLTPWMDYEFVFIASNLGTG 694
 Qy 635 VPSGLSNVKTIEADFAASANDLSAARTLLTGKSVELIDASAINASAVRLEWMLHVSAD 694
 Db 695 EPSGSSKIRTREA-APSVAPSGLSGG-----GGAPGELI---VNWTPMSREYQ---NGD 742
 Qy 695 E-KYVEGLR----IHYKDVSVPSAQYHSITVMDASAESFVVG--LKYYTKYEFFLTPTFF 747
 Db 743 GFGYLLSFRQGSTHWQTVARVPG-----ADAQYFVYSNESVRYTPPEVKIRSY- 791
 Qy 748 ETIEGQPSNSKALTLY--EDVPSAPPDNIQIMYNTAGVWRWTPPPSQHNGNLYGYKI 805
 Db 792 -NRGDCGPESLTALVYSAEEPRVAPTQVWAKGVSSSEMNVTW-EPVQDMGILLGYEI 849
 Qy 806 EV-SAGNTMKVLANMTLNATTSVLLNNLTGAVISVRLNSFTKAGDGPYSKPI-SLFMD 863
 Db 850 RWKAGDKAAADRVRTAGLDTARSVGLHPNTKTHVTVRAYNRAGTGPASPSANATMK 909
 Qy 864 PTHHVHPRAHPSG-----THDGRHEGQDLTYHNNGNIPPGDINP 903
 Db 910 P-----PPRPPGNISWTFSSSSLSIKWDPVVPFRNESAVTYGKMLYQNDLH-----LTP 959
 Qy 904 TTHKKTIDYLSGP 916
 Db 960 TLHLTGKNWIEIP 972

RESULT 8

CAML_MOUSE

ID CAML_MOUSE STANDARD; PRT; 1260 AA.

AC P11627;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).

GN L1CAM OR CAML1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=BRAIN;

RX MEDLINE=88318924; PubMed=3412448;

RA Moos M., Tacke R., Scherer H., Teplow D., Frueh K., Schachner M.;

RT "Neural adhesion molecule L1 as a member of the immunoglobulin

RT superfamily with binding domains similar to fibronectin.";

RL Nature 334:701-703(1988).

CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE

CC DEVELOPMENT-OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
 CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
 CC TO AXONIN ON NEURONS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
 CC PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X12875; CAA31368.1; -.
 DR PIR; S05479; S05479.
 DR HSSP; P20241; 1CFB.
 DR MGD; MGI:96721; L1CAM.
 DR INTERPRO; IPR001777; -.
 DR INTERPRO; IPR003006; -.
 DR PFAM; PF00041; fn3; 4.
 DR PFAM; PF00047; ig; 6.
 DR PRINTS; PR00014; FNTYPEIII.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
 FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1124 1146 POTENTIAL.
 FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 346 410 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 440 503 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 531 599 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 827 896 FIBRONECTIN TYPE-III.
 FT DOMAIN 932 994 FIBRONECTIN TYPE-III.
 FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III.
 FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1180 1183 MISSING IN SHORT ISOFORM
 FT (BY SIMILARITY).
 SQ SEQUENCE 1260 AA; 140968 MW; 22BE57001CB2A538 CRC64;

Query Match 8.6%; Score 641.5; DB 1; Length 1260;
 Best Local Similarity 23.0%; Pred. No. 7.4e-25;
 Matches 275; Conservative 167; Mismatches 458; Indels 295; Gaps 48;

Qy 35 WLLLVLASNGLPAVRGQYQSPRIE-----HPTDLVVKNEPATLNCKVEGKPEPTI 87
| | | : | : : | : | | : | : | : | :
Db 9 WPLL-LCSPCLLIQIPDEYKGGHVLPEPVITEQSPRRLLVFPDDISLCKEARGRQVEF 67

Qy 88 EWFKDGEVPSTNEKKS---HRVQFKDGLFFYRTMQGKK---EQDGGYWCVAKNRVGQA 141
| | | : | : : | : | : | : | : | : | :
Db 68 RWTKDGIIHFKPEELGVVVEAPY-SGSF---TIEGNSFAQRFGIYRCIASNKLGA 122

Qy 142 VSRHASLIQAVLRDDFRVEPKDT---RVAKGETALLECGPKGIPEPTLIWIKDGV--- 194
: | : | : : | : | : | : | : | : | :
Db 123 MSH---EIQLVAEGAPKWKETVKPVEVEEGESVVLPCNPPPSAAPPRIYWMNSKIFI 178

Qy 195 -----PLD-----DLK 200
| : | : | : | : | : | : | : | :
Db 179 KQDERVSMGONGDLYFANVLTSDNHSYICNAHFPGTRTIQKEPIDLRVKPTNSMIDRK 238

Qy 201 ---AMSFAGSSRVIVDGGNLLIS----- 221
: | | : | : | : | : | : | :
Db 239 PRLFFPTNSSRLVALQGSLLILECIAEGFPTPTIKWLHPSDPMPTDRVIYQNNHKLQL 298

Qy 222 -NVEPIDEGNYKICIAQLVGTRESSYAKLIVQKPYFMKEPKDQVMLYGGTATPHCSVGG 280
| | : | : | : | : | : | : | : | : | : | :
Db 299 LNVGEEDGGEYTCLENSLGSARHAY-YVTVEAAPPYWLKQKSHLYGPGETARLDCCVQG 357

Qy 281 DPPPKVLWKKEGNIPVSARILHDEK-----SLEISNITPTDEGTYVCEAHNNVGGIS 334
| : | : | : | : | : | : | : | : | : | :
Db 358 RPQPEITWRIN---GMSMETVNDQKRYEQGSLLSNVQPTDMVTQCEARNQHGLLL 413

Qy 335 ARASL-IVHAPPNFKRPSNKKVGLNG-VVQLPCMASGNPPSVFWTKEGVSTLMFPNSS 392
| : | : | : | : | : | : | : | : | : | :
Db 414 ANAYIYVQLPARILTKDNQTYMAVEGSTAYLLCKAFGAPVPSQWLDEEGTYL---Q 469

Qy 393 HGRQYVAADGTLQITDVRQEDGYYVCSAFSVSDSSTVRVFLQVSSVDERPPPIIIGPA 452
| : | : | : | : | : | : | : | : | : | :
Db 470 DERFFPYANGTSLIRDQANDTGRYFCQAANDQNNVTILANLQVKEATQ-----ITQGP 524

Qy 453 NQTLPGKSVATLPCRATGNPS--PRIKWFHDGHAFA---GNRYSIQSSSLRVDDLQLS 507
| : | : | : | : | : | : | : | : | : | :
Db 525 SAIEKKGARVTFQASFDPSLQASITWRGDRDLQERGEDSKY-FIEDGKLVIQSLDYS 583

Qy 508 DSGTYICTASGERGET-SWAATLTVEKPGSTSLHRAADPSTYPAPPGTPKVLNVSTIS 566
| : | : | : | : | : | : | : | : | : | :
Db 584 DQGNYSVASTELDEVESRAQLLVGSGPGVPHLESDRHL-----LKQSQVH 631

Qy 567 LRWAKSQEKPGAVGPIIGTYEYFSPDL-QTGWIVAAHRVGDQTQVITISGLTPGTSYVFLV 625
| : : : | : : : | : : : | : : : | : : : | : : :
Db 632 LSWSPAEDHN---SPIEKYDIEPEDKEMAPEKWFSLGKVPNGQSTTLKLSPYVHYTFRV 688

Qy 626 RAENTQGISVPSGLSNVIKTIEADFAASANDLSAA-----RTLLTGKSVELIDASAI 678
| | | | : | : | : | : | : | : | : | :
Db 689 TAINKYGCEPSPVSESVVTPPEA---APEKNPVDVRGEGNETNMVITWKLRMWMDNAP 745

Qy 679 NASAVRLEWMLHVSADKTYEGLRIHYKSDASVPSAQYHSITVMDASAFVVGNLKYYTK 738
| : | : | : | : | : | : | : | : | : | :
Db 746 QIQ-YRVQWR-----PQKQETWRKQTVSDPFLVVSNTSTFVP 782

Qy 739 YEFFLTPFFETIEGQPSNSKTALTY---EDVPSAPPNDIQIGMYNTAGWVRWTPPPSQ 794
| | : : : | : | : | : | : | : | : | :
Db 783 YEIKV---QAVNNQKGPEPQVTIGYSGEDYQVSPLELIDITIFNSSTVLVWRVPDLA 838

Qy 795 HHNGNLYGYKI-----EVSAGNTMKVLANMTLNATTSVLLNNLTGAVSVRLNS 845
| : | : | : | : | : | : | : | : | : | :
Db 839 QVKGHLKGYNVTYWKGSRQKHSKRHIK--SHIVVPANTTSAILSLGRPYSSHYVEQA 896

Qy 846 PTKAGDGP-----YSKPISLMDPTH-----HVHPRAHPSGTHDG--- 881
| | | : | : | : | : | : | : | : | :
Db 897 FNGRGLGPASEWTFSTPEGV---PGHPEALHLEQSDTSLLEHWQPLSH-NGVLTGYLL 952

Qy 882 -RH--EGQ-----DLTYHNNGNIPPGDINTTHKKTDTYLSGFWMLVLCIVL 926
| | : : | : | : | : | : | : | : | : | :
Db 953 SYHPVEGESKEQLFNLSDEPLRTHNLNLNP-DLQYRFLQATTTQGGGPEAIVREGGT 1011

Qy 927 LVLV-----ISA-----AISMVYFK-----RKHQMTKELGHLSSVSDNEITALNIN 967
: | | : | : | : | : | : | : | : | :
Db 1012 MALFGKPDFGNISATAGENYSVSVWVRKQCNRFFHILFKALPEKGVSPDHQPQPOQYVS 1071

Qy 968 SKESLWIDHRRGWR-TADTDKDSGLSESKLLSHVNSQSNYNNSDGGTDYAEVDT 1021
: | : : | : | : | : | : | : | : | :
Db 1072 YNQSSTQ---WNLPDQTKYEHILKEKVLHLLDVKTN-----GTGPVRVST 1116

RESULT 9

DCC_MOUSE

ID DCC_MOUSE STANDARD; PRT: 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RX MEDLINE=96112625; PubMed=8570174;
RA Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
gene (mDCC) and its expression in the developing mouse embryo."
RL Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC AND REMAIN AT THIS LEVEL IN THE ADULT.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X85788; CAA59786.1; -.
DR HSSP; P56276; 1TLK.
DR MGD; MGI:94869; DCC.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, LONG
FT ISOFORM.
FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, SHORT
FT ISOFORM.
FT INIT_MET 85. 85 FOR SHORT ISOFORM.

FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1098 1122 POTENTIAL.
FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 54 124 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 154 219 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 317 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 819 838 MISSING (IN EMBRYONIC ISOFORM).
SQ SEQUENCE 1447 AA; 158298 MW; 0D1F1097C22D5B9F CRC64;

Query Match 8.6%; Score 639.5; DB 1; Length 1447;
Best Local Similarity 23.1%; Pred. No. 1.1e-24;
Matches 272; Conservative 151; Mismatches 453; Indels 303; Gaps 43;

Qy 53 YQSPRIIEHPTDLVVKNEPATLNCKVEG-KPEPTIEWKDG--EPVSTNEKSHRVQFK 109
Db 37 FTSLHFVSEPSDAVTMRGGNVLLNCSAESDRGVPIKWKDGLILALGMDDRKQ---QLP 93
Qy 110 DGALFFYRTMQGKKEQ-DGGEYWCVAK-NRVGQAVSRHASLQIA-VLRDDFRVEPKDTRV 166
Db 94 NGSLLIQNILHSRHHKPDGLYQCEASLADSGSIISRTAKVTAGPLR--FLSQTESITA 151
Qy 167 AKGETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNNLLISNVEPI 226
Db 152 FMGDTVLLKC-EVIGEMPPTIHWQKNQDLNPLP-----GDSRVVVLPSGALQISRLQPG 205
Qy 227 DEGNKYCIAQNLVGTRESSYAKLIVQVKP-----YFMKEPKDQVMVYGQTATFHC SVGG 280
Db 206 DSGVYRCSARNPASIRTGNAEVRILSDPLGHRQLYFLQRPSNVIAEGDKAVLECCVSG 265
Qy 281 DPPPKVLWKKEGNIPV-SRARILHDEKSLISNITPTDEGTYVCEAHNVGQISARSL 339
Db 266 YPPPSFTWLRGEEVIQLRSKKYSLGGSNLLISNVTDDSGTTCVVTYKNENISASAE 325
Qy 340 IVHAPPNFTKRPSNKKVGLNGVQLPCMASGNPPSVFVTKEGVSTLMPNSSHGRQYVA 399
Db 326 TVLVPPWFLNHPSN-----LYATESM----- 346
Qy 400 ADGTLQITDVRQEDEGYVCSAFSVSDSSTVRVFLQVSSVDERPPPIIQIGPANQLPKG 459
Db 347 -----DIEFF----- 351
Qy 460 SVATLPCRTAGNPSRIKWFHDGHAVQAGNYSIIQGSLSRVDLQLSDSGTYTCTASGE 519
Db 352 -----CAVSGKPVPTVNMKNGDVVPSDYQIVGGSNLRILGVVKSDEGYPQCAENE 405
Qy 520 RGETSWAATLTVEKPGSTSLHRAADPSTYPAPPPTPKVLNVSRISLWRKASQEKPGAV 579
Db 406 AGNAQSSAQLIVPKPAIPS-----SSILPSAPRDVPLVLSRFRVLSWRPPAE---AK 456
Qy 580 GPIIGTYVEYFSPD-----LQTGWIAAHRVGDQTQVITSGLTPTGTSYVFLVRAENTQG 632
Db 457 GNIQTFTV-FFSREGDNRRALNT-----TQPSGLQLTVGNLKPAMTYFRVAYNEWG 509
Qy 633 ISVPSGLSNVIKTIEADFAASANDLSAARTLLTGKSVELIDASAINASAVRLEWLHVS 692

Db 510 ---PGESSQPIK-----VATQPELQVPGPVENLHAVSTSPSTSLITWEPPAY 553
Qy 693 ADEKYVEGLRIHYKDSVPSAQYHSITVMDASAESEFVGNLKKYTKYEFFLTPFFETIEG 752
Db 554 ANG-PVQGYRLCTEVSTGKEQN-----IEVDGLSYKLEGLKKFTEYTLRFLAYNRVGP 607
Qy 753 QPSNSKTALTYEDVPSAPPDNIQIGMYNQTAGWVRWTPPPSQHNGNLYGYKI---EVSA 809
Db 608 VSTDITVVTLSDVPSAPPQNISLEVNVNSRIKSVLPPPSGTQNGFTTGYKIRHKTRT 667
Qy 810 GNTMKVLNMTLNATTISVLLNNLTGAVISVRLNSFTKAGDGPYSPKISLFMDPTHVHV 869
Db 668 RGE ME-----TLEPNLWYLTGKLEKGSQYSQVSAMTVNGTGP----- 706
Qy 870 PPAHPSGTHDRHEGQDLTYHNNNIP--PGD--INPTTHKKTIDYLSGPWMLVLCIV 925
Db 707 -----PSNWTAEPTENDL--DESQVDPQSSLHVRPQTN-----CII 742
Qy 926 LLVLVISAASIMVYFKRKHQMTKELGHLVSDNEITALNINS--KESLWIDHHRGWRTA 983
Db 743 M-----SWTPLL-NPNIVVRGYIIGYGVGSPTAETVRVDSQRYSI 783
Qy 984 DTDKDSGLSESKLLSHVNSSQSNYNNSDGGTD-YAEVDTRNLTTFYNCRSPDNPTPYAT 1042
Db 784 ERLSS-----SHYVISLKAFFNAGEGVPLYESATRSIT-----DPTDPVDYYP 828
Qy 1043 TMIIGTSSSETCTKTTISADKDSGTHSPYSDAFAGQVPAVPVVKSNLYQYVPEINWSE 1102
Db 829 LL-----DDFPTSGP--DVSTPMLPPVG-VQAVALTAEAVRVSWAD 866
Qy 1103 FLPPPEHPPPSSSTYGYAQGSPSSRSKSSAGSGISTNQSIILNASIHSSSGGFSAWGV 1162
Db 867 NSVPKNQKTSIVRLTVRWRTSFSASAKYKS-----EDTSLSYTATGL 910
Qy 1163 SP-----QYAVACPPENVNINPLSAVAGGTQNRQYITPTN 1197
Db 911 KPNTMYEFSVMV-TKNRRSSTWSMTAHAT--TYEAAPTS 946

RESULT 10

DCC_HUMAN

ID DCC_HUMAN STANDARD; PRT; 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011532; PubMed=7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RT tumorigenesis."
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE=90100559; PubMed=2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers."
RL Science 247:49-56(1990).
RN [3]
RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX MEDLINE=91121517; PubMed=1991322;
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,

RA Oliner J.D., Kinzler K.W., Vogelstein B.;
RT "Scrambled exons";
RL Cell 64:607-613(1991).
RN [4]
RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX MEDLINE=94245241; PubMed=8188295;
RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT "The DCC gene: structural analysis and mutations in colorectal
RT carcinomas";
RL Genomics 19:525-531(1994).
RN [5]
RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX MEDLINE=94243823; PubMed=8187090;
RA Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT human esophageal squamous cell carcinomas and their relation to
RT metastasis";
L Cancer Res. 54:3007-3010(1994).
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC -1- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X76132; CAA53735.1; -.
DR EMBL; M32292; AAA35751.1; -.
DR EMBL; M32286; AAA52174.1; -.
DR EMBL; M32288; AAA52175.1; ALT_SEQ.
DR EMBL; M32290; AAA52176.1; -.
DR EMBL; M63696; AAA52177.1; -.
DR EMBL; M63700; AAA52178.1; -.
DR EMBL; M63702; AAA52179.1; -.
DR EMBL; M63718; AAA52180.1; -.
DR EMBL; M63698; AAA52181.1; -.
DR PIR; A54100; A54100.
DR PIR; A40098; A40098.
DR PIR; A38442; A38442.
DR HSSP; P56276; TLK.
DR MIM; 120470; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Anti-oncogene; Disease mutation; Polymorphism.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC.
FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1098 1122 POTENTIAL.
FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 54 124 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 154 219 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 317 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III.

FT DOMAIN 722 816 FIBRONECTIN TYPE-III.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 168 168 M -> T (IN OESOPHAGEAL CARCINOMA).
FT /FTid=VAR_003909.
FT VARIANT 201 201 R -> G.
FT /FTid=VAR_003910.
FT VARIANT 1375 1375 P -> H (IN A COLORECTAL CARCINOMA).
FT /FTid=VAR_003911.
FT CONFLICT 138 138 MISSING (IN REF. 3).
FT CONFLICT 233 329 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
SQ SEQUENCE 1447 AA; 158456 MW; 4A861276ED0471F CRC64;

Query Match 8.6%; Score 637.5; DB 1; Length 1447;
Best Local Similarity 22.9%; Pred. No. 1.4e-24;
Matches 271; Conservative 150; Mismatches 444; Indels 317; Gaps 43;

Qy 57 RIIEHPTDLVVKNEPATLNCKVEG-KPEPTIEWFKDG--EPVSTNEKSHRVQFKDGL 113
Db 41 RFLSEPSDAVTMRGGNVLLDCSAESDRGVPVVKWKGDIHLALGMDERKQ---QLSNGSL 97
Qy 114 FFYRTMQGKKEQ-DGGEYCVAK-NRVQAVSRHASLQIA-VLRDDFRVEPKDTRVAKGE 170
Db 98 LIQNLHLSRHHKPDGLYQCEASLGDSGSIIISRTAKVAVAGPLR--FLSQTESVTAFMGD 155
Qy 171 TALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGFSSRRVIVDGGNLLISNVEPIDEGN 230
Db 156 TVLLKC-EVIGEMPMTIHWQKNQDLTPIP-----GDSRVVLPSPALQISRLQPGDIGI 209
Qy 231 KYCIAQNLVGTRESSYAKLIVQVKP-----YFMKEPKQVMLYGTATFCHSVGGDPPP 284
Db 210 YRCSARNPASSRTGNEAEVRILSDPGLHRQLYFLRPSNVVAIEGKDAVLECCVSGYPPP 269
Qy 285 KVLWKEEGNIPV-SRARILHDEKSLISNITPTDEGTYYCAHNNVGQISARASLIVHA 343
Db 270 SFTWLGRGEEVIQLRSKKYSLGGSNLLISNVTDDSGMYTCVITYKNENISASAEALTVLV 329
Qy 344 PPNFTKRPSNKKVGLNGVVQLPCMASGNPPSVFVTKEGVSTLMPNSSHGRQYVAADGT 403
Db 330 PPWFLNHPSN-----LYAYESM----- 346
Qy 404 LQITDVRQDEBGYYVCSAFSVVDSSTVRVFLQVSSVDERPPPIIQIGPANQTLPGKSVAT 463
Db 347 ----DIEFE----- 351
Qy 464 LPCRATGNPSPRIKWFHDGHAVQAGNRYIIQGSLLVDDQLSLSGTYTCTASGERGET 523
Db 352 --CTVSGKPVPTVNMKNMGDVPVPSDYFQIVGGSNLRILGVVKSDEGYPQCAVNEAGNA 409
Qy 524 SWAATLTVKEKPGSTSLHRAADPSTYPAPPGTPKVLNVRSISLRWAKSQEKPAGVPII 583
Db 410 QTSQQLIVPKPAIPS-----SSVLPSAPRDVVPVLVSSRFVRLSWRPPAE---AKNGIQ 460
Qy 584 GYTVVEYFSPD-----LQTGWIVAHRVGDQTQVTSGLTPGTSYVFLVRAENTQGISVP 636
Db 461 TPTV-FFSREGDNRRALNT-----TQPGSLQLTVGNLKPAMYTFRRVAVNEWG---P 510
Qy 637 SGLSNVIKTIADFDAAASANDLSAARTLLTCKSVELIDASAINASAVRLEWMLHVSADK 696
Db 511 GESSQPIK-----VATQPELQVPGPVENLQAVSTSPSTILITWEPPAYANGP 557

RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, L1.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]
RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in L1-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; V-439--T-443 DEL; C-784 AND L-936--L-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).
RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE=98112489; PubMed=9452110;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
Hum. Mutat. Suppl. 1:S284-S287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE=98415726; PubMed=9744477;
RA Saugier-Verber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
RT "Identification of novel L1CAM mutations using fluorescence-assisted
RT mismatch analysis.";
RL Hum. Mutat. 12:259-266(1998).
CC #1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF THREE X-LINKED
CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
CC SYLVIVUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
CC Milder PRESENTATION AND A LONGER LIFE EXPECTANCY. 3: SPASTIC
CC PARAPLEGIA TYPE 1 (SPG1). COLLECTIVELY THESE SYNDROMES ARE ALSO
CC KNOWN AS CRASH SYNDROME, AN ACRONYM WHICH STANDS FOR CORPUS
CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,

CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF HIRSCHPRUNG DISEASE
CC (HSCR).
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=L1CAM; NOTE=L1CAM mutation Web Page;
CC WWW="http://hgins.uia.ac.be/dnalab/l1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59847; CAA42508.1; -.
DR EMBL; Z29373; CAA82564.1; -.
DR EMBL; M74387; AAA59476.1; -.
DR EMBL; X58775; CAA41576.1; -.

Query Match 8.5%; Score 633.5; DB 1; Length 1257;
Best Local Similarity 23.7%; Pred. No. 1.8e-24;
Matches 243; Conservative 142; Mismatches 395; Indels 245; Gaps 36;

Qy 35 WLLLVLVASNLPAVRGQYQSPRIIE-----HPTDLVVKKNPATLNCNKVEGKPEPTI 87
| | | | : : : : | | | | : : : : | | | | : : : : | | | |
Db 9 WPLL-LCSPCLLIQIPPEYEGHHVMEPPVITEQSPRLVVFPTDDISLCKEASGKPEVQF 67
Qy 88 EWFKDGEPVSTNEKSHRVQFKDGLFFRYTMQGGK----EQDGGEYWCVAKNRVQAVS 143
: : | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 68 RWRTRDGVHFKPEELGVTVYQSPHSGSF--TITGNNSNFAQRFQGIYRCFASNKLGATMS 125
Qy 144 RHASLIQIAVLRRDFRVEPKDT----RVAKGETALLECGPPKGIPEPTLIWIKDGVPLDDL 199
: | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 126 H---EIRLMAEGAPKPKETVKPVEVEEGESVVLPCNPPPSAEPLRIYWMNS----- 174
Qy 200 KAMSFAGSRRVIVDGGNLLISNV-----EPI----- 226
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 175 KILHIKQDERVTMGQNGNLYFANVLTSNHSYICAHFPGTRTIIQKEPIDLRVKATNS 234
Qy 227 ----- 226
Db 235 MIDRKPRLLFPTNSSSHLVALQGQPLVLEICIAEGFPTPTIKWLRPSGPMADRVTYQNNH 294
Qy 227 -----DEGNKICIAQNLVGTRESSYAKRLIVQVKPYFMKEPKDQVMLIGQTATFH 275
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 295 KTLQLLVGEGDDGEYRCLAENSIGSARHAY-IVTVEAAPYWLHKQSHLYGPGGTARLD 353
Qy 276 CSVGGDPPPKVLWKKEGNIPIVSRARILHDEK-----SLEISNITPDEGTIVCEAHNN 329
| | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 354 CQVQGRPQPEVTVR--INGIPVE--ELAKDQYRIQKALILSNVQPSDTMTVQCEARNR 409
Qy 330 VQGISARSL-IVHAPPNFTKRPSNKKVGLNG-VVQLPCMASGNPPPSVFWTKGVSTLM 387
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 410 HGLLLANATIVVQLPAKILTADNQYMAVQGSTAYLLCKAFGAPVPSVQWLEDGTTVL 469
Qy 388 FPNSSHGQYVAADGTLQITDVRQEGDEGYVCSAFSVSDSTTVRVFLQVSSVDERPPPII 447
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 470 ---QDERFFPYANGTLGIRDQLANDTGRYFCLAAQNNVTIMANLKVKDQATQ-----I 520
Qy 448 QIGPANQTLPAQSVATLPCRATGNPS--PRIKWFHDGHAVQA---GNRYSIIGSSLRVD 502
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 521 TQGPRTIEKKGSRVFTFCQASFPDQLSPSITWRGDRDLQELGSDSKY-FIEDGRLVH 579
Qy 503 DLQLSDSGTYICTASGERGET-SWAATLTVEKPGSTSLHRAADPSTYPAPPPTKVLN-V 560
| | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 580 SLDSYDQGNYSVASTELDVVDSRAQLLVVSGPG-----PVPRVLVSDHLHL 626
Qy 561 SRTSISLRWAKSQEKPGAVGPIIGYTVVEYFSPDL-QTGWIVAAHVRGDTQVTISGLTPTG 619
: : : : : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 627 TQSQVRVWSNPAEDHN---APIEKYDIEFEDKEMAPEKWSLGVKPGNQSTTLKLSPTV 683

```

Qy 105 WLLVLVLVASNG--LPAVRGQYQSPRIE-----HPTDLVVKKNEPATLNCKVEGKPEP 85
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 106 WYVLPDLLLCSPELLIQIPDEYKGHVHEPPVITEQSPRRLVVFPTDDISLKCEARGRPQV 65

Qy 186 TIEWFKDGEVPYSTNEKKS---HRVQFKDGLFFYRTMQGKK---EQDGGEYWCVAKNRVG 139
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 187 EFRRTKDGIFHKPEELGVVVHEAPY-SGSF-----TIEGNNSTFAQRFGQIYRCYASNNG 120

Qy 310 QAVSRHASL-----QIAVLRD 155
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 311 TAMSHEIQLVAEGAPKPKETVKPVEVEEGESVVLPCNPPPSAAPLRIYWMNSKILHIQ 180

Qy 456 DERV-----EPKDRV-----166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 457 DERVSMGQNGDLYFANVLTSDNHSDYICNAHFGPRTTIQKEPIDLRVKPTNSMIDRKPR 240

Qy 617 -----AKGETALLECGPKGIPEPTLIWI--KDGVPDLLKAMSFSGASSRV 111
   : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 618 LLFPTNSSSHLVALQGQSLLILEC-IAEGFPTPTIKWLHPSDPMPTD-----R 286

Qy 712 IV---DGGNLLISNVEPIDEGNYKCIAQLNVGTRESSYAKLIVQKVPYFMKEPKDQVMYL 268
   : : | : | | : | | | : | : | : | : | : | : | : | : | : | :
Db 713 VIYQNHNTKLQLNNGEEDDEYTCIAENSLGSAKHAY-YTVTEAAPYWLQKPQSHLYGP 345

Qy 829 GQTATFHCSVGGDPPPKVLWKKEEGNIPVSRARILHDEK-----SLEISNITPTDEGTY 322
   | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 830 GETARLDCOVGRPOPEVTTWRIN---GMSIEKVNKDOKYRIEGBSLLSNVQPSDPTMVT 401

```

Qy 323 VCEAHNNVGQISARASL-IVHAPPNFTKRPNSKKVGLNG-VVQLPCMASGNPPSVFWTK 380
| | | | : : : : | : : : : | | | | | | |
Db 402 QCEARNQHGILLANAYIYVQVLPARILKDNQTMVAVGSEYALLCKAFGAPVPSVQWLD 461

Qy 381 EGVSTLMPNSSHGRQYVAADGTLQITDVRQDEGYVCSAFSVDSSTVRVFLQVSSVD 440
| : : : | : : | | : : | | : : | : : | : : | : : |
Db 462 EGGTVL-----QDERFPFYANGHLGIRLDQANDTGRYFCQAANDQNNVTILANLQVKEAT 517

Qy 441 ERPPPIIIGPANQTLPGKSVATLPCRATGNPS--PRIKWFHDGHAVQA---GNRYSIQ 495
: | : : | : : | : : | : : | : : | : : | : : | : : |
Db 518 Q-----ITQGPSTIEKKGARVFTTCQASFDPSLQASITWRGDRDLQERGDSKY-FIE 571

Qy 496 GSSLRVDDLQSDSGTYTCTASGERGET-SWAATLTVEKPGSTSLHRAADPSTYPAPPGT 554
| : : | | | | | | | | | | | | | : |
Db 572 DGQLVKSLDYSDQGDYSCVASTELDEVESRAQLLVVSGPGVPVHLESLDRHL----- 624

Qy 555 PKVLNVSRTISLRWAKSQEKPAGVPIIGTYVEYFSPDL-QTGWIVAHRVGTQVTTIS 613
: : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 625 -----LKQSQVHLSWSPAEDHN---SPIEKYDIEFEDKEMAPEKWFSLGKVPGNQSTTL 676

Qy 614 GLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIEADFAASANDLSAA-----RTLLT 666
| : | | | | | | | | | | | | | : : | : : |
Db 677 KLSPIYVHTFRVTAINKYGGEPSVSETVVTPEA---APEKNPVDVRGEGNETNMVIT 733

Qy 667 GKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDSVPSAQYHSITVMDASAE 726
| : : : | : : | : : | : : | : : | : : | : : | : : |
Db 734 WKPLRWMDWNPQIQ-YRVQWR-----PLGKQETWKEQTVSDP 770

Qy 727 SFVGNLKKYTYEFLTPFFETIEGQPSNSKTALTY----EDVPSAPPDNIQIGMYNQ 782
| | : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 771 FLVVSNTSTFPYIEKV---QAVNNQKGPPEQVITIGYSGEDYQVSPLEDTITFNSS 826

Qy 783 AGWVRWTPPPSQHNGNLYGYKI-----EVSAGNTMKVLNMTLNATTTSVLLNNL 833
| | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 827 TVLVRRVVDLAQVKGHLRGYNVTYWWKGSQRKSKRHVHK---SHMVVPANTTSAILSG 884

Qy 834 TTGAVYSVRLNSFTKAGDGP-----YSKPISLFMDPH-----HVHPPR 872
: | : : | : : | : : | : : | : : | : : | : : | : : |
Db 885 RPYSSHYEVQAFNGRGLGPASEWTFSTPEGV---PGHPEALHLEQSDTSLLLHWQPL 941

Qy 873 AHPS-----GTH--DGRHEGQ-----DLTYHNNGNIPPG-----DINPTHK 907
: | : : | : : | : : | : : | : : | : : | : : | : : |
Db 942 SHNGVLTGYLLSYHPLDGESKEQLFFNLSPDLRTHNLNLPDLQYRFQLQATTHQ 998

SULT 13

01_CHICK
ID AX01_CHICK STANDARD; PRT; 1036 AA.
AC P28685;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-BRAIN;
RX MEDLINE=92174898; PubMed=1311675;
RA Zuellig R.A., Rader C., Schroeder A., Kalousek M.B.,
RA von Böhlen Und Halbach F., Osterwalder T., Inan C., Stoeckli E.T.,
RA Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
RT "The axonally secreted cell adhesion molecule, axonin-1. Primary
RT structure, immunoglobulin-like and fibronectin-type-III-like domains
RT and glycosyl-phosphatidylinositol anchorage."
RL Eur. J. Biochem. 204:453-463(1992).
CC -!- FUNCTION: AXON-ASSOCIATED CELL ADHESION MOLECULE (AXCAM) WHICH
CC PROMOTES NEURITE OUTGROWTH BY INTERACTION WITH THE AXCAM L1 (G4)
CC OF NEURITIC MEMBRANE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A

CC GPI-ANCHOR.
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63101; CAA44815.1; -.
DR PIR; S22128; S22128.
DR PIR; S22383; S22383.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 23 OR 25 (POTENTIAL).
FT CHAIN 24 1036 AXONIN-1.
FT PROPEP ? 1036 REMOVED IN MATURE FORM.
FT DOMAIN 49 113 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 143 211 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 249 308 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 336 397 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 428 490 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 518 589 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 599 608 HINGE (POTENTIAL).
FT DOMAIN 601 607 GLY/PRO-RICH.
FT DOMAIN 608 709 FIBRONECTIN TYPE-III.
FT DOMAIN 710 811 FIBRONECTIN TYPE-III.
FT DOMAIN 812 912 FIBRONECTIN TYPE-III.
FT DOMAIN 913 1009 FIBRONECTIN TYPE-III.
FT MOD_RES 724 724 BLOCKED.
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1036 AA; 113301 MW; 08B80143BE779794 CRC64;

Query Match 8.4%; Score 626.5; DB 1; Length 1036;
Best Local Similarity 24.2%; Pred. No. 3.2e-24;
Matches 241; Conservative 130; Mismatches 390; Indels 233; Gaps 35;

Qy 71 EPATLNCKVEGKPEPTIEWFKDGEPVSTNEKKSRRVQFKDGLAFFYRTMQGK-----K 123
| | | : : | | | : : | | | : : |
Db 50 EKVTLTCRARANPPATYRW-----KMGTELKMGPDSTRYLAVAGDLVISPVK 97

Qy 124 EQDGGEYWCVAKNRVQAVSRHASLQIAVLRRDFRVEPKD-TRVAKGETALLECGPPKGI 182
: | | | | | | | | | : : | : : | : : | : : |
Db 98 AKDAGSYQCVATNARGTVVSREASLRFGLQ-EFSABERDPVKITEGWMFTCSPPHY 156

Qy 183 PEPTLIWIKDGVPLDDLKAMSGASSRVRIVD--GGNLLISNVEPIDEGNYKICIAQLVG 240
| : : : | : : | : : | | | : : | | | | : : |
Db 157 PALSYRWLLNEFP-----NFIPADGRRFVSQTGNLYIAKTEASDLIGNYSCFATSHID 209

Qy 241 --TRE--SSYAKLIV-----QVKPYF-MKEPKDQVMYLGQTATFHCVGGDPPPKVLWK 289
: | : : : : | : : | | | | | | : : : : | : : : : |
Db 210 FTTKSVFSKFSQLSLAEDARQYAPSIKAFPADTYALTGMVTECFAGFNPVQIKWR 269

Qy 290 KEEGNIPVSARILHDEKSLISNITPTDEGTYVCEAHNNVGQISARASLIVHAPPNFTK 349
| : : : : | : : | | | | | | : : : : | : : : : |
Db 270 KLDGS---QTSKWLSEPLLHIQNVDFDEGTYCEAENIKGRDTYQGRIIHAQPDWLD 326

Qy 350 RPSNKKVGLNGVQLPCMASGNPPPSVFWTKEG-----VSTLMFPNS- 391
Db 327 VITDTEADIGSDLRWSCVASKPRPAVFWLRDGPPLASQNRIEVSGGELRFSKLVLEDSG 386
Qy 392 -----SHGRQY----- 397
Db 387 MYQCVAKNGHGTVYASAEITVQALAPDFRLNPVKRLIPAARSKGVIIPQAPRAKATVL 446
Qy 398 -----VAADGTLQITDVRQDEGEYVCSAFSVVDSSTVRVFLQVSSVDER 442
Db 447 WTKGTELLTNSSRVITADGTLILQNKISDEGKYTCFAENPMKANSTGILSV-----R 501
Qy 443 PPPIIQIGPANQTLPGKSVATLPCRATGNPSPI--KWFHD-----GHAQVQAGNR 490
Db 502 DATKITLAPSSADINVENLTLQCHASHDPTMDLFTWSLDDFPIDLDKSEGHYRRASVK 561
Qy 491 YSIQGSLSRLVDDLQSDSGTYTCTASGERGETSWAATLTVEKPGSTSLHRAADPSTYPA 550
Db 562 EAV---GDLAIVNAQLKHSGRYTCTAQTVDSTSESATLTVRGP-----PG 604
Qy 551 PPGTPKVLNVRSISLRWAKSQEKGAVGPIIGTYVEYFSPDLQGWIAAHRVGDQV 610
Db 605 PPGGVVVRDIDGTTVQLSWGRGDNH---SPIARYSIEARTL-LSNKK---KQMRNPV 656
Qy 611 TISG-----LTPGTSYVFLVRAENTQGISVPSGLSNVIKTIADFDAAASANDLSAA 661
Db 657 NIEGNAETAQVNVNLIPLMDYEFRLASNLGVEPSLSPSKIRTKEA-APTAPSGGLGGG 715
Qy 662 RTLLTGKSVELIDASAINASAVRLEWM-----LHVSADKTYEGLRIHYKDVSPSAQY 715
Db 716 ---GGAPNELI---INWPTLRLDYQNGDGFYILSFKKGTQC---WLARVPHAE- 762
Qy 716 HSITVMDASAESFVVGNN--LKKYTKYEFFLTFFPETIEGQPSNSKALTYY--EDVPSAPP 771
Db 763 -----SLHYVYRNESIGPYTPEVKIKAY--NRKGEGPSLTAIVISAEEEPKVP 811
Qy 772 DNIQIGMYNQTAGWVRWTPPPSQHNGNLYGKIEV-SAGNTMKVLNMTLNATTSVLL 830
Db 812 FRVTAKAVLSSEMDVSWVEVEQDGMTGLVGLYEIRYKWDGDKKEAADRVTRAGLVTSAHV 871
Qy 831 NNLTTGAVYSVRLNSFTKAGDGPYSKPSLFMDPTHVHVPPRAHPSGTHDRHGQDLTY 890
Db 872 TGLNPNTKYHVSRYNRRAGAGPPSPSNTITT-----TKPPRRPPGNISWTLTGSTVTI 926
Qy 891 HNNGNIPPGDINPTT-----HKKTTDYLS 914
Db 927 KWDPVVAQADESAVTGKMLYRQDSHSAPTLYLA 960

RESULT 14

NRCA_CHICK

ID NRCA_CHICK STANDARD; PRT; 1284 AA.
AC P35331;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NG-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE=91258407; PubMed=2045418;
RA Grumet M., Mauro V., Burgoon M.P., Edelman G.M., Cunningham B.A.;
RT "Structure of a new nervous system glycoprotein, Nr-CAM, and its
relationship to subgroups of neural cell adhesion molecules."
RL J. Cell Biol. 113:1399-1412(1991).
RN [2]
RP SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-EMBRYONIC BRAIN, AND RETINA;

RA MEDLINE=92381110; PubMed=1512296;
RA Kayyem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;
RT "Bravo/Nr-CAM is closely related to the cell adhesion molecules L1
and Ng-CAM and has a similar heterodimer structure."
RL J. Cell Biol. 118:1259-1270(1992).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC. SPECIFICALLY INVOLVED IN THE DEVELOPMENT OF OPTIC
CC FIBRES IN THE RETINA.
CC -1- SUBUNIT: HETERODIMER, COMPOSED OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RETINA AND DEVELOPING BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING NEURAL RETINA AND
CC EMBRYONIC BRAIN TISSUE.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58482; CAA41391.1; -.
DR EMBL; L08960; AAA48632.1; -.
DR HSSP; P20241; 1CFB.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 5.
DR PFAM; PF00047; ig; 6.
DR PRINTS; PR00014; FNTYPEIII.
KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW Transmembrane; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 1284 NG-CAM RELATED CELL ADHESION MOLECULE.
FT DOMAIN 25 1143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1144 1166 POTENTIAL.
FT DOMAIN 1167 1284 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 56 125 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 155 220 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 261 323 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 351 415 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 445 508 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 599 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 638 699 FIBRONECTIN TYPE-III.
FT DOMAIN 738 799 FIBRONECTIN TYPE-III.
FT DOMAIN 837 906 FIBRONECTIN TYPE-III.
FT DOMAIN 943 1006 FIBRONECTIN TYPE-III.
FT DOMAIN 1057 1114 FIBRONECTIN TYPE-III.
FT DISULFID 63 118 POTENTIAL.
FT DISULFID 162 213 POTENTIAL.
FT DISULFID 268 316 POTENTIAL.
FT DISULFID 358 408 POTENTIAL.
FT DISULFID 452 501 POTENTIAL.
FT DISULFID 543 592 POTENTIAL.
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

Db      567 ELLIRNAQLKHAGRYTCTAQTIVDNSSASADLVVRGPPGGRIEDIRATSVALTWSR 626
Qy      572 SQEKPGAVGPIIGYTV-----YFSPDLQGTGWIAAAHRVGDVTISGLTPGT 619
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      627 GSDNH---SPI SKYTITQTKILSDDWKDAKDTPPIIEGNMEARAV-----DLIPWM 675
Qy      620 SYVFLVRAENTQGISVP SGLSNVIKTIEADFDAASAN---DLSAARTLLTGKSVELIDA 675
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      676 EYEFRVVATNLTGRGEPSIPSNRIKT----DGAAPNVAPSVDGCG--GGRNREL--- 723
Qy      676 SAINASAVRLWMLHVSADKEYEGLRIHYKASVP--SAQYHSITVMDASAESFVVGN- 732
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      724 -----TITW---APLSREYHYGNFGYI AFAKPFDEGEWKVTVTNPDTGRIVHKDE 772
Qy      733 -LKKYTKYEFFLTPFFETIEGQPNSKALTIEDVPSAPPDNIQIMYNQTAGVWRWTPP 791
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      773 TMSPTAFQKVYKAFNNKGDPISLVAVINSAQDAPSEAPTGVGKVLSSSEISVHW--- 829
Qy      792 PSQHNGNLVGYKIEVSAGNTMKVLAN-MTLNATTSVLLNLLTGVAVYSVRLNSFTKA 849
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      830 --EHVLEKI VESYQIRYWAAHDKEEAANRVQVTSQEYSARLENLLPDQTYFIEVGACNSA 887
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Qy      850 GDGPYSKPISLFMDPTHVHPPPRA-----HPS 876
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      888 GCGPPSDMI EAFTKKAPPSQPRIISSVRSGRSYITWDHVVALSNESTVTGYKVLVRPD 947
Qy      877 GTHDGR 882
        : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      948 GOHDGK 953

```

Query Match 8.3%; Score 614.5; DB 1; Length 1018;
Best Local Similarity 22.8%; Pred. No. 1.3e-23;
Matches 234; Conservative 143; Mismatches 404; Indels 245; Gaps 35;

Search completed: January 22, 2001, 12:27:18
Job time: 1139 sec

```

Qy 28 RMWLLPALLLLVLVAS-----NGLPAVRGQYQSPRIIEHPTDLVVKKNE---P 72
   :||| :|:: : : :| :| : :
Db 2 KMWLLVSHLVIISITCLAEFTWYRRYGHGV-SEEDXGFGPIFEQPIINTIYPEESLEG 60

Qy 73 ATLCKVEGKPEPTIEWFKDGEVPVSTNEKSHRVQFKDGLFFYRTMGKKQDGGEYTC 132
   :||| :| :| :| :| :| :| :| :| :| :|
Db 61 VSLNCRARASPFPPYKWRMNGNDV---DLTSDRYSMVGGNLVINNP--DKQKDAGIYTC 114

Qy 133 VAKNRVGQAVSRHASLQIAVLRDDFRVEPK-DTRVAKGETALLECGPPKGIPEP-TLIWI 190
   :| | | :| :| :| :| :| :| :| :| :| :|
Db 115 LASNNYGMVRSATEATLSFGYL-DPPFPEERPEVRKVEKGMVLLCDPPYHFPDDLSTRWL 173

Qy 191 KDGVPDDLKAMSGASSRVRIVD--GGNLLISNVEPDEBGNKYCIAQNLVGTRESSYAK 248
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 174 LNEFPV-----FITMDKRRFVSQTNGNLYIANVEASDKGNYSCFVSSPSITK-SVFSK 225

Qy 249 LIVQV-----KPY--FMKEPKQVMLYGQTATFHCSVGGDPPPKVLWKKEEGNIPVS 298
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 226 FIPLIPIPERTTKYPADIVVQFKDVALYMGQNVTLCEFALGNVPVDIRWRKVLPM-P-S 284

Qy 299 RARILHDEKSLEISNITPTDEGTYVCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVGL 358
   | | :| :| :| :| :| :| :| :| :| :| :|
Db 285 TAEISTSGAVLKIPNIQLEDEGIYCEAENIRGDKKHQARIYVQAPPEWHEHINDTEVDI 344

Qy 359 NGGVQLPCMASGNPPPSVFWTKEGVSTLMPNSSHGRQYVAADGTLQITDVROQDEGYTV 418
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 345 GSDLYPWCVATGKPIPTIRWLKNG-----YAHKGELRLYDVTFENAGMTQ 390

Qy 419 CSAFSVSDSSTVRVFLQVSSVDERPPPIIQIGPANQTL--PKGSVATLPCRATGNPSPRI 476
   | | : : :| : : :| : : :| : : :| : : :|
Db 391 CIAENYGAIIYANAELKILAL---APTFFMNMKKKILAAKGGRVIECKPKAAPKPKF 446

Qy 477 KWFHDGHAVQAGNRYSIQSSSLVRDDLQLSDSGTYTCTASGERGETSWAATLTVPEK-- 534
   | : :| :| :| :| :| :| :| :| :| :| :|
Db 447 SWSKGTWLVNSSRRLIWEDEGSLEINNITRNDGGIYTCFAENNRKANSTGLTIVITDPT 506

Qy 535 -----GSTSLHRAA---DPS--TY----- 548
   | : :| :| :| :| :| :| :| :| :| :| :|
Db 507 IILAPINADITVGENATMQCAASFDPALDLTFVWSFNGYVIDFNKENIHQRFNMLDSNG 566

Qy 549 -----PAPPGTKPVLNVSRTSISLRWAK 571

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:50:23 ; Search time 559.88 Seconds
(without alignments)
292.036 Million cell updates/sec

Title: US-09-540-245A-15
Perfect score: 7427
Sequence: 1 MHPMPENHAIARSTSTNN.....SCLYAEAGEPAPQMTAKNT 1395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	7427	100.0	1395	5	Q44924	Q44924 drosophila
2	7407	99.7	1395	5	Q9W213	Q9W213 drosophila
3	1620	21.8	859	5	Q9VP26	Q9VP26 drosophila
4	1609.5	21.7	1273	5	Q44928	Q44928 caenorhabdi
5	1607.5	21.6	1612	11	O89026	O89026 mus musculus
6	1592	21.4	1651	4	Q9Y6N7	Q9Y6N7 homo sapien
7	1585	21.3	1651	11	O55005	O55005 rattus norv
8	1499	20.2	1060	11	Q9QZ13	Q9QZ13 rattus norv
9	1430	19.3	823	5	Q9VQ10	Q9VQ10 drosophila
10	1395.5	18.8	1344	11	Q9Z214	Q9Z214 mus musculus
11	826.5	11.1	423	5	P91572	P91572 caenorhabdi
12	790	10.6	874	5	O01632	O01632 caenorhabdi
13	759.5	10.2	1377	11	P97603	P97603 rattus norv
14	731.5	9.8	1493	11	P97798	P97798 mus musculus
15	712.5	9.6	1461	4	O00340	O00340 homo sapien
16	710.5	9.6	1461	4	Q92859	Q92859 homo sapien
17	686	9.2	1277	13	Q98902	Q98902 fugu rubrip
18	677	9.1	1443	13	Q90610	Q90610 gallus gall
19	667.5	9.0	1822	4	Q9UL77	Q9UL77 homo sapien

20	663.5	8.9	1026	4	O94780	O94780 homo sapien
21	659.5	8.9	1100	4	O94779	O94779 homo sapien
22	652	8.8	2016	5	Q9V4J9	Q9V4J9 drosophila
23	651	8.8	1026	11	Q62845	Q62845 rattus norv
24	651	8.8	2221	5	Q9ULM1	Q9ULM1 drosophila
25	649	8.7	2016	5	Q9NBA1	Q9NBA1 drosophila
26	647	8.7	1028	11	Q62682	Q62682 rattus norv
27	645.5	8.7	1445	11	Q63155	Q63155 rattus norv
28	645	8.7	1028	11	Q07409	Q07409 mus musculus
29	644	8.7	1272	13	Q90924	Q90924 gallus gall
30	643	8.7	1217	11	P97685	P97685 rattus norv
31	636	8.6	2222	5	O97394	O97394 drosophila
32	635	8.5	1099	11	P97527	P97527 rattus norv
33	632	8.5	1264	5	P91767	P91767 manduca sex
34	631	8.5	1369	13	Q42414	Q42414 gallus gall
35	629	8.5	1259	11	Q9QY38	Q9QY38 mus musculus
36	628	8.5	1028	11	P97528	P97528 rattus norv
37	626.5	8.4	1248	6	Q9XT41	Q9XT41 cercopithec
38	624	8.4	1028	11	Q9JMB8	Q9JMB8 mus musculus
39	622.5	8.4	1427	13	Q91562	Q91562 xenopus lae
40	619.5	8.3	1018	6	Q28106	Q28106 bos taurus
41	616.5	8.3	1302	5	O61542	O61542 drosophila
42	616	8.3	1166	11	Q9QVM4	Q9QVM4 rattus sp.
43	612.5	8.2	1151	11	Q9QVN5	Q9QVN5 rattus sp.
44	612.5	8.2	1215	11	P97686	P97686 rattus norv
45	611	8.2	1239	5	Q9V3X0	Q9V3X0 drosophila

ALIGNMENTS

RESULT 1
O44924
ID O44924 PRELIMINARY; PRT; 1395 AA.
AC O44924;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040989; AAC38849.1; -.
DR HSSP; P56276; IYLK.
DR FLYBASE; FBgn0035631; robo.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 1395 AA; 151778 MW; B820E234A5218983 CRC64;

Query Match 100.0%; Score 7427; DB 5; Length 1395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHPMPENHAIARSTSTNNPSRSRSSRWLLPAWLLLVASNGLPAVRGQYQSPRIIE 60

Db 1 MHPMPENHAIARSTSTNNPSRSRSSRWLLPAWLLLVASNGLPAVRGQYQSPRIIE 60

Qy 61 HPTDLVVKKNPATLCKVEGKPEPTIEWFKDGEVSTNEKSHRVQFKDGLFFYRTMQ 120

Db 61 HPTDLVVKNEPATLNCKVEGKPEPTIEWFKDGEVPSTNEKSHRVQKFDGALFFRTMQ 120
Qy 121 GKKEQDGGGYWCYVAKNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Db 121 GKKEQDGGGYWCYVAKNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Qy 181 GIPEPTLIWIKDGVPLDDLKAMSGASSRVIVDGNLLISNVEPIDEGNYKICIAQLVVG 240
Db 181 GIPEPTLIWIKDGVPLDDLKAMSGASSRVIVDGNLLISNVEPIDEGNYKICIAQLVVG 240
Qy 241 TRESSYAKLIVQVQPYFMKEPKDQVMLYQGTATFHC SVGGDPPPKVLWKKKEGNIPVSR 300
Db 241 TRESSYAKLIVQVQPYFMKEPKDQVMLYQGTATFHC SVGGDPPPKVLWKKKEGNIPVSR 300
Qy 301 RILHDEKSLISNITPTDEGTYVCEAHNVGQISARASLIVHAPPNFKRPSNKKVGLNG 360
Db 301 RILHDEKSLISNITPTDEGTYVCEAHNVGQISARASLIVHAPPNFKRPSNKKVGLNG 360
Qy 361 VVQLPCMASGNPPSVFWTKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQEDGEGYVCS 420
Db 361 VVQLPCMASGNPPSVFWTKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQEDGEGYVCS 420
Qy 421 AFSVVDSTVRVFLQVSSVDERPPPIIQIGPANQTLPGKSVAILPCRAAGNPSRIKWFH 480
Db 421 AFSVVDSTVRVFLQVSSVDERPPPIIQIGPANQTLPGKSVAILPCRAAGNPSRIKWFH 480
Qy 481 DGHAVQAGNRYSIQGSLLRVDLQLSDSGTYTCTASGERGETSWAATLTVEKPGSTSLH 540
Db 481 DGHAVQAGNRYSIQGSLLRVDLQLSDSGTYTCTASGERGETSWAATLTVEKPGSTSLH 540
Qy 541 RAADPSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGTYVEYFSPDLQGTGIV 600
Db 541 RAADPSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGTYVEYFSPDLQGTGIV 600
Qy 601 AAHRVGDQTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSA 660
Db 601 AAHRVGDQTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSA 660
Qy 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDAVSQAQYHSITV 720
Db 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDAVSQAQYHSITV 720
Qy 721 MDASAESFVVGNLKYYTYEFLTPFFETIEGQPSNSKALTIEDVPSAPPDNIQIGMYN 780
Db 721 MDASAESFVVGNLKYYTYEFLTPFFETIEGQPSNSKALTIEDVPSAPPDNIQIGMYN 780
Qy 781 QTAGVVRWTPPPSQHNGNLGYKIEVSAGNTMKVLANMTLNATTSVLLNLTGAVYS 840
Db 781 QTAGVVRWTPPPSQHNGNLGYKIEVSAGNTMKVLANMTLNATTSVLLNLTGAVYS 840
Qy 841 VRLNSFTKAGDGPYSKPISLFMDPTHVHPRAHPSGTHDGRHEGQDLTYHNGNIPPGD 900
Db 841 VRLNSFTKAGDGPYSKPISLFMDPTHVHPRAHPSGTHDGRHEGQDLTYHNGNIPPGD 900
Qy 901 INPTTHKKTIDYLSGFWMLVLCIVLLVLVISAAMVYFKRKHMTELGHLSVVS DNE 960
Db 901 INPTTHKKTIDYLSGFWMLVLCIVLLVLVISAAMVYFKRKHMTELGHLSVVS DNE 960
Qy 961 ITALNINSKESLWIDHHRGWRTADTKDGLSGLSKLLSHVNSQSNYNSDGGTDYAEVD 1020
Db 961 ITALNINSKESLWIDHHRGWRTADTKDGLSGLSKLLSHVNSQSNYNSDGGTDYAEVD 1020
Qy 1021 TRNLTTFFNCRKSPDNPTPIATTMIIGTSSSETCTKTTISADKDSGTHSPYS DAFAGV 1080
Db 1021 TRNLTTFFNCRKSPDNPTPIATTMIIGTSSSETCTKTTISADKDSGTHSPYS DAFAGV 1080
Qy 1081 PAVPVVKSNYLQYVPEPINWSEFLPPPEHPPPSSTYGYAQSPESSRKS SAGSIST 1140
Db 1081 PAVPVVKSNYLQYVPEPINWSEFLPPPEHPPPSSTYGYAQSPESSRKS SAGSIST 1140
Qy 1141 NQSILNASIHSSSSGGFSAWVSPQYAVACPPENVSNPLSAVAGGTQNRQYIIP TNPQH 1200

Db 1141 NQSILNASIHSSSSGGFSAWVSPQYAVACPPENVSNPLSAVAGGTQNRQYIIP TNPQH 1200
Qy 1201 PQLPAYFATGPGGAVPPNHLFPATQRHAASEYQAGLNAARCAQSRACNSCDALATPSM 1260
Db 1201 PQLPAYFATGPGGAVPPNHLFPATQRHAASEYQAGLNAARCAQSRACNSCDALATPSM 1260
Qy 1261 QPPPPVPVPGWYQPVHPNSHPMHTSSNHQIYQCSSECDHSRSSQSHKRLQLEEHGS 1320
Db 1261 QPPPPVPVPGWYQPVHPNSHPMHTSSNHQIYQCSSECDHSRSSQSHKRLQLEEHGS 1320
Qy 1321 SAKQGGHRRRAPVVPQCMESENENMLAEYEQRYTSDCCNSSREGDTCSSEGSCLYA 1380
Db 1321 SAKQGGHRRRAPVVPQCMESENENMLAEYEQRYTSDCCNSSREGDTCSSEGSCLYA 1380
Qy 1381 EAGEPAPRQMTAKNT 1395
Db 1381 EAGEPAPRQMTAKNT 1395

RESULT 2

Q9W213

ID Q9W213 PRELIMINARY; PRT; 1395 AA.

AC Q9W213;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE ROBO PROTEIN.

GN ROBO.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003458; AAP46887.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0005631; robo.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; PNTYPEIII.
SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13F0 CRC64;

Query Match 99.7%; Score 7407; DB 5; Length 1395;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1391; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MHPMPENHAIARSTSTNNPSRSRSMWLLPAWLLLVASNGLPAVRGQYQSPRIE 60
Db 1 MHPMPENHAIARSTSTNNPSRSRSMWLLPAWLLLVASNGLPAVRGQYQSPRIE 60
Qy 61 HPTDLVVKNEPATLNCKVEGKPEPTIEWFKDGEVPSTNEKSHRVQFKDGLFFYRTMQ 120
Db 61 HPTDLVVKNEPATLNCKVEGKPEPTIEWFKDGEVPSTNEKSHRVQFKDGLFFYRTMQ 120
Qy 121 GKKEQDGGYWCYVAKNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Db 121 GKKEQDGGYWCYVAKNRVQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETALLECGPPK 180
Qy 181 GIPEPTLIWIKDGVPLDDLKAMSGASSRVIRVDGNGLLISNVEPIDEGNYKCIANLVG 240
Db 181 GIPEPTLIWIKDGVPLDDLKAMSGASSRVIRVDGNGLLISNVEPIDEGNYKCIANLVG 240
Qy 241 TRESSYAKLIVQKPYFMKEPKDQVLMYQGTATFHC SVGGDPPPKVLWKKEGNIPIVSR 300
Db 241 TRESSYAKLIVQKPYFMKEPKDQVLMYQGTATFHC SVGGDPPPKVLWKKEGNIPIVSR 300
Qy 301 RILHDEKSLAISNITPTDEGTYVCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVGLNG 360
Db 301 RILHDEKSLAISNITPTDEGTYVCEAHNNVGQISARASLIVHAPPNFTKRPSNKKVGLNG 360
Qy 361 VVQLPCMASGNPPSVFWTKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQDEGYVCS 420
Db 361 VVQLPCMASGNPPSVFWTKEGVSTLMFPNSSHGRQYVAADGTLQITDVRQDEGYVCS 420
Qy 421 AFSVDSSTVRVFLQVSSVDERPPPIIQIGPANQLPKGSVATLPCRATGNPSRIKWFH 480
Db 421 AFSVDSSTVRVFLQVSSVDERPPPIIQIGPANQLPKGSVATLPCRATGNPSRIKWFH 480
Qy 481 DGHAVQAGNRYSIQGSLLRVDDLQLSDSGTYTCTASGERGETSWAATLIVKPGSTSLH 540
Db 481 DGHAVQAGNRYSIQGSLLRVDDLQLSDSGTYTCTASGERGETSWAATLIVKPGSTSLH 540
Qy 541 RAADPSTYPAPPPTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGTYVEYFSPDLQGTGWIV 600
Db 541 RAADPSTYPAPPPTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGTYVEYFSPDLQGTGWIV 600
Qy 601 AARVGDGTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSA 660
Db 601 AARVGDGTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVKTIEADFDAASANDLSA 660
Qy 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDASVPSAQHSITV 720
Db 661 ARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIHYKDASVPSAQHSITV 720
Qy 721 MDASAESFVVGNLKKYTKYEFFLTPTFFETIEGQPSNSKTALTIEDVPSAPPDNIQIGMYN 780
Db 721 MDASAESFVVGNLKKYTKYEFFLTPTFFETIEGQPSNSKTALTIEDVPSAPPDNIQIGMYN 780
Qy 781 QTAGWVRWTPPPSQHHNGNLYGKIEVSAGNTMKVLNMTLNATTTSVLLNLTGAVIS 840
Db 781 QTAGWVRWTPPPSQHHNGNLYGKIEVSAGNTMKVLNMTLNATTTSVLLNLTGAVIS 840

Qy 841 VRLNSFTKAGDGPYSKIPISLMDPTHVHPPRAHPSGTHDGRHEGQDLTYHNNGNIPPGD 900
Db 841 VRLNSFTKAGDGPYSKIPISLMDPTHVHPPRAHPSGTHDGRHEGQDLTYHNNGNIPPGD 900
Qy 901 INPTTHKKTIDYLSGPWMLVLCIVLLVISAASIMVYFKRKHQMTKELGHLVSVSDNE 960
Db 901 INPTTHKKTIDYLSGPWMLVLCIVLLVISAASIMVYFKRKHQMTKELGHLVSVSDNE 960
Qy 961 ITALNINSKESLWIDHHRGWRTADTDKDSGLSESKLLSHVNSSQSNYNNSDGGTDYAEVD 1020
Db 961 ITALNINSKESLWIDHHRGWRTADTDKDSGLSESKLLSHVNSSQSNYNNSDGGTDYAEVD 1020
Qy 1021 TRNLTTFPYCNCRSPDNPTPYATTMIIGTSSETCTKTTISADKDSGTHSPYSDAFAGQV 1080
Db 1021 TRNLTTFPYCNCRSPDNPTPYATTMIIGTSSETCTKTTISADKDSGTHSPYSDAFAGQV 1080
Qy 1081 PAVPVVKSNIYLPVPEPINWSEFLPPPEHPSPSTYGYAQSGPSSSRKSSKAGSGIST 1140
Db 1081 PAVPVVKSNIYLPVPEPINWSEFLPPPEHPSPSTYGYAQSGPSSSRKSSKAGSGIST 1140
Qy 1141 NQSIILNASHSSSSGGFSAWGVSQYAVACPPENVISNPLSAVAGGTQNRQITPTNQHP 1200
Db 1141 NQSIILNASHSSSSGGFSAWGVSQYAVACPPENVISNPLSAVAGGTQNRQITPTNQHP 1200
Qy 1201 POLPAYFATTGPGGAVPPNHLFPFATQRAHASEYQAGLNAARCAQSRACNSDALATPSM 1260
Db 1201 POLPAYFATTGPGGAVPPNHLFPFATQRAHASEYQAGLNAARCAQSRACNSDALATPSM 1260
Qy 1261 QPPPPVPVPEGWYQVHPNHPMHTSSNHQIYQCSSECDHSRSSQSHKRQLQLEEHGS 1320
Db 1261 QPPPPVPVPEGWYQVHPNHPMHTSSNHQIYQCSSECDHSRSSQSHKRQLQLEEHGS 1320
Qy 1321 SAKORGHRRRAPVVQPCMESEENMLAEYEQRYTSDCCNSSREGDTSCSEGSCLYA 1380
Db 1321 SAKORGHRRRAPVVQPCMESEENMLAEYEQRYTSDCCNSSREGDTSCSEGSCLYA 1380
Qy 1381 EAGEPAPQMTAKNT 1395
Db 1381 EAGEPAPQMTAKNT 1395

RESULT 3

Q9VPZ6

ID Q9VPZ6 PRELIMINARY; PRT; 859 AA.

AC Q9VPZ6;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CG5423 PROTEIN (FRAGMENT).

GN CG5423.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; *Drosophila*.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

Qy 156 PRIIEHPTDLVVKKNKEPATLNCKVVEGKPEPTIEWFKDGEVPVSTNEKSHRVQFKDGLFF 115
   |||:| | | :||| | | | | | | | | | | | : | | | | | | |
Db 1 PRIVEHPDITTVPRHEPATLNCKAEGSPITPTIQWKDGVPLKI-LPGSHRITLPAGGLFF 59

Qy 116 YRTMQGKKEQDGGEYWCVAKNRVQAVSRHASLQIAVLRRDDEFVEPKDTRVAKGETALLE 175
   : ||: | | | | | | | | | | | | | | | | | | | | | |
Db 60 LKVSDDGR-----CAVLRFDEFLEPQNTRIAQGDUTALLE 93

Qy 176 CGPPKGIPEPTLIWIKDGVPLDLKAMSFGASSRVRIVDGGNLLISNVEPIDEGNYKCIA 235
   | :|:| | | | | | | | | | | | | | | | | | | | | |
Db 94 CAAPRGIPETPTVTKKGQKLD-----LEGSKRRVRIVDGGNLAIQDARQTDGQYQICA 147

Qy 236 QNLVGTRESSYAKLIVQKVPYFMKEPKDQVMLYGQTATFHCSVGGDPPPKVLWKK--EEG 293
   : | | | | | | | | | | | | | | | | | | | | | |
Db 148 KNPVGVRESSLALKVHVKPYIIRGPHDQTVLEGASVTFPCRVGGDMPDVLWLRTASGG 207

Qy 294 NIPVSRARILHDEKSLEISNITPTDEGTYYCEAHNNVQGISARASLIVHAPPNFTKRPSN 353
   | : | : | : | : | : | | | | | | | | | | | | | : | : | :
Db 208 NMPLDRVSVLED-RSLRLERVITADEGEYSCEADNVVGAITAMGTLTYAPPKFIQRPAS 266

Qy 354 KKVGLNGVVQLPCMASGNPPPSVFVTKEGVSTLMPNSS-----HGRQYVAADGTLQITD 408
   : | : | : | : | : | | | | | | | | | | | | | | | |
Db 267 KSVELGADTSFECRAIGNPKPTFTIKNSTLIFPGAPPLDRFHSLSNTEEGHSILTLTR 326

Qy 409 VRQEDEGYIV-CSAFSVVSDSTVRVFLQVSSVDERPPPIIQIGPANQTLPGKSVATLPCR 467
   : | : | : | | | | | : | : | | | | | | | | | | | : | : |
Db 327 FQRTDKDLVILCNAMNEVASITSRVQLSLDSQEDRPPPIISGPNVQTLPIKSLATLQCK 386

Qy 468 ATGNPSPRIKWFHDGHAVQAGNRYISIQGSSLRVDDL-QLSDSGTYTCTASGERGETSWA 526
   | | | | | | | | | | | : | : | | | | | | | | | | | : | : |
Db 387 AIGLPSPTISWYRGIPVQSSKLNITTSGLDIISDLRQDQGLYTCVASSRACKSTWS 446

Qy 527 ATLTVKPGSTSL--HRAADPSTYPAPPGTKVLNVSRITSISLRWAKSQEKGAVGPIIG 584
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 4
O44928
ID O44928 PRELIMINARY; PRT; 1273 AA.
AC O44928;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE SAX-3.
GN SAX-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID:6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98117250; PubMed-9458046;
RA Zallen J.A., Yi B.A., Bargmann C.I.;
RT "The conserved immunoglobulin superfamily member SAX-3/Robo directs
RT multiple aspects of axon guidance in C. elegans.";
RL Cell 92:217-227(1998).
DR EMBL; AF041053; AAC38848.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNYPEIII.
SQ SEQUENCE 1273 AA; 139427 MW; 013E766B51A7BAD7 CRC64;

```

```
Qy 55 SPRRIIEHPTDVLVKKNEPATLNCKVEGKPEPT-IEWFKDGEVPSTNEKK--SHRVQFKDG 111
   :| ||||| |:|: : ||||| | | |:|:| | |:|: |||: |
Db 30 APVIEEHPIDVVSVRGSPATLNC--GAKPSTAKITWYKDGQPVITNKEQVNSHRIVLDTG 87

Qy 112 ALFFYRTMQGK--KEQDGGEYVCWKNRVGQAVSRHASLQIAVLRRDFRVEPKDTRVAKG 169
   :| |||: | |:| :| |||: | | | | | | | | | | | | | | | |
Db 88 SFLFLKYSNGKNGKSDAGAYCYVANSNEHGEVKSNEGSLKLAMLRDFRVRPTVOALG 147
```

```

Qy      56 PRIIEHPTDLVVKKNPATLNCKVEGKPEPTIEWFKDGEVPST---NEKSKSHRVQKFDGAL 113
      |||:| |:| |:| | ||||| | |:| |||||:| || | | :: :|||: | |:|
Db      29 PRIVEHPSDLIVSKGEPATLNCKAEGRPTPTIEWKGGGERVETDKDDPRSHRMLPSGSL 95

Qy     114 FFYRTMQGKKEQ-DGGEYWCVAKNRVGQAVSRHASLQIAVLRDDFRVEPKDTRVAKGETA 172
      || | : |:| : | | |||:| |:| ||| :|||:|:| |||| | | || | |
Db      89 FFLRIVHGKSRPDEBGYVICVARNYLGEAVSHNASLEVALLRDDFRQNPSDVMVAVGEP 148

Qy     173 LLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFAGSSRVRIVDGGNLLISNVPEIDEGNYK 232
      :|| |||:| |||:| | ||| ||| | :| | || |:| | | |
Db     149 VMECQPPRGHPEPTLISWKKDGSPDD-----KDERITI-RGKKLMTITYTRKSDACKTV 200

Qy     233 CIAQNLVGTRESSYAKLIVQVKPYFMKEPKDQVMLYQGTATFHC SVGDDPPPKVLWKKEE 292
      |: |:| || || |:| |:| |:| |:| : | |:| |:| || |:| |:| :|
Db     201 CVGTNMVGERSSVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDD 260

Qy     293 GNIPVSRARIHDEKSLEISNITPTDEGYVCEAHNNVVGISARASLIVHAPPNFTKRPS 352
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     261 GELPKSRYEI-RDDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTVQEPHPHVVKPR 319

Qy     353 NKKVGLNGVQLPCMASGNPPPSVFVTKEGVSLMF---PNSSHGQRQYVAADGTLQITDV 409
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     320 DQVVALGRTVIFQCEATGNPQPAIFWRREGSQNLLFSYQPPQSSSRFSVSQTGDLTITNV 379

Qy     410 RQEDEGYIVCSAFSVVSDSTVRVFLQVSSV-DERPPPIIQIGPANQTLPGKSVATLPCRA 468
      |:| |:| |:| | | | |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     380 ORSDVGYIICOTNLVAGSIITKAYLEVTDVIADRPVPIVROGPVNOTVAVDGTLILSCVA 439

```


Db 713 ----PSGANHGESDVLFEVVRTPAKNSVVPDLRKGVNTEIKARPFNEFGADSEIKFA 768
Qy 761 LTYEDVPSAPPDNIQIGMY--NQTAGWVRWTPPPSQHHNGLYGYKIEVSAGNTMKVLAN 818
| : | | | | : : | | | | | : | : | : | : |
Db 769 KTLLEAPSPAPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQYEVK-WCLGNETRYHIN 827
Qy 819 MTLNATTTISVLLNLTGAVISVRLNSFTKAGDGPYSKISLFMDPTHVHPRAHPSGT 878
| : : | | : : | | | : : | | : : | : |
Db 828 KTVDGSTFSVVPFLVPGIRYSVEVAASTGAGSGVKSEPFQIQLD-----AH---- 874
Qy 879 HDGRHEGQDLTYHNGN-IPPGDINPTTHKKTIDYLSGP-----WLMVLVCIVLL 927
| : | : | : : : : | : : : | : : : |
Db 875 -----GNPVSPEL-QVSLAQISDVVKQPAFIAGIAGAACWILMVFSIWL 918
Qy 928 VLVISAAISMVYFKRKHQ-----MTKELGHLSSVSDNEITALNI 966
| | : : : : : : : : : : : : | | |
Db 919 -----YRHKRRNGLTSTYAGIRKVPSTFTPTVYQRGGEAVSSGGRPGLLNI 967
967 NKSESL-WI-----DHRGWRTADTKDGLSGLSKILLSHVNSQ--SNYNNS----- 1010
: : : : : : : : : : : : : : : : : :
Db 968 SEPAAQPLWADTWPNTGNHNDSCISCTAGNGNSDNLTTYSRPACIANYNQDLNKO 1027
Qy 1011 -----DGGTDYAEVDTRNLTTFYNCRKSPP-----NPTPYATMIIGTSSSETC 1054
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1028 TNMLPESTVYGDVLSNKNEMKTFNSPNLKDGRFVNPSPQPTPYATTLQIQLNSNM 1087
Qy 1055 TKTTISADK-----DSGTH 1069
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1088 NNGSGDSGEKHWKPLGQKQEVAPVQYNIQVNEQNKLNDYRANDTVPTPIPNQSYDQNTG 1147
Qy 1070 SPYSDAFAGQVPAVPVVKSNYLQYVPEP----INWSEFLPPPEHPPPSSTGYAGGSPE 1125
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1148 GSYNSDGRGSGTSGSQGHKKGARTPKVPKQGMWADLPPPPAHPPHS----- 1197
Qy 1126 SSRKSSKSAGSGISTNQSLNASIHSSSSGGFSANGVSPQYAVACPPENVY----- 1176
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1198 -----NSEEYINISVDES-----YDQEMPCVPPARMYQQDELEEE 1233
Qy 1177 -----SNPLSAVAGGTQNRQYQITPTNQH--PPQLPAYFATTGPGGAVPPNH 1220
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1234 EDERGTPPPVGAASSP-AAVSYSHQSTAILTPSQEELQMLQDCPEETG-----H 1284
Qy 1221 LPFATQRHAASEYQAGLNAARCAQSRACNSCDALATPSMPQPPPPVPEGWYQVPHNS 1280
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1285 MQHQPDRRR-----QPVSPPPP-PRP-----ISPPTYG 1312
1281 HPMHPTSSN 1289
: : | : :
Db 1313 YISGPLVSD 132L

RESULT 7
O55005
ID O55005 PRELIMINARY; PRT; 1651 AA.
AC O55005;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TRANSMEMBRANE RECEPTOR ROBO1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF041082; AAC39960.1; -.

DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
KW Transmembrane.
SQ SEQUENCE 1651 AA; 180746 MW; FA2452DD46E186B7 CRC64;

Query Match 21.3%; Score 1585; DB 11; Length 1651;
Best Local Similarity 29.0%; Pred. No. 5.1e-98;
Matches 456; Conservative 207; Mismatches 538; Indels 374; Gaps 49

Qy 44 NGLPA-----VRGQYQSPRIIEHPTDLVVKNEPATLNCKVEGKPEPTI 87
| | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 40 NCTPAPTSDNDDNSLGYTGSRLRQEDFPFPRIVEHPSDLIVSKGEPATLNCKAEGRPPTI 99
Qy 88 EWFKDGPEVST--NEKKSRRVQFKDGFALFFYRTMQGKKEQ-DGGEYWCVAKNRVGQAVSR 144
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 100 EWFKGGERVETDKDDPRSHRMLPSGSLFFLRIVHGRKSRPDEGVICVARNYLGEAVSH 159
Qy 145 HASLIQAVLRDDFRVEPKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPLDLKAMSF 204
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 160 NASLEVAIRLDDFRQNPDSVMVAVGEPVMECPRGHPPTISWKKDGSPLD----- 213
Qy 205 GASSRVIVDGGNLLISNVEPIDEGNYKCIQNLVGTRESSYAKLIVQVKPYFMKEPKDQ 264
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 214 -KDERITI-RGKLMITYTRKSDAGYVCVGTMMVGERESKVADVTVLERPSFVKRPSNL 271
Qy 265 VMLYGTATFHCSSVGGDPPPKVLWKEEGNIPVSRARILHDEKSLISNITPTIDEGTYVC 324
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 272 AVTVDDSAEFKCEARGDPVPTFGWRKDDGELPKSYEI-RDDHTLKIRKVTAGDMGSIYC 330
Qy 325 EAHNNVQISARASLIHVAPPNFTKRPSNKKVGLNGVVLPCMASGNPPPSFWTKEGVS 384
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 331 VAEHMYGKAESATLTQEPHPFVVKPRDQVVALGRVTVFQCEATGNQPAIFWRREGSQ 390
Qy 385 TLMF---PNSSHGRQYVAADGTLQITDVRQDEGYVCSAFSVSDSTVRVFLQVSS-D 440
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 391 NLLFSYQPPQSSSRFSVSQTGDLVTNVQRSDVGYICQLTNVAGSIITKAYLEVTDVIA 450
Qy 441 ERPPPIIQIGPANQTLPGKSVATLPCRATGNPSRIKWFHDGHAVQA-GNYSIIQSSSL 499
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 451 DRPPVPVIRQGPVNQTVAVDGTLTLSCVATGSPVPTILWRKDGVLVSTQDSRIKQLESQVL 510
Qy 500 RVDDLQSDSSTYTCTASGERGETSWAATLTVEKPG-STSLHRAADPSTYPAPPGTPKVL 558
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 511 QIRYAKLGDGTCTASTPSGEATWSAYIEVQEPGVPPQPRPTDPLNIPSAPSKEPTV 570
Qy 559 NVSRTISLWRKWSQEKPGAVGPIIGITYVEYFSPDLQGWIVAAHRVGDQVTISGLTPG 618
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 571 DVSKNTVTLW---QPNLNSGATPTSIIIEAFSHAGSSSWQVAENVKTETFAIKGLKN 627
Qy 619 TSYVFLVRAENTQGISVPSGLSNVIKTIEADFAASANDLSAARTLLTGKSV-ELIDASA 677
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 628 AIYFLVRAANAYGISDPSQISDPVKTQDPPTQGVVDHQQVQREL--GNVVLHLHNPTI 685
Qy 678 INASAVRLEWMLHVSADKYEVEGLRIHYK--DASVPSAQYHSITVMDASAEFVGNLKK 735
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 686 LSSSSVEVHWI--VDQQSQYIQGYKILYRPSGASHGESEWLVFEVVRTPTKNSVVPDLRK 743
Qy 736 YTKYEFLTPFFETIEGQPSNKALTIEDVPSAPPDNIQIGMY--NQTAGWVRWTPPPS 793
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 744 GVNTEIKARPFNEFGADSEIKFAKTLERPSAPPSPVTVSKNDGNGTAILVTPQPPPE 803
Qy 794 QHHNGLYGYKIEVSAGNTMKVLANMTLNTATTSVLLNLTGAVISVRLNSFTKAGDGP 853
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 804 DTQNGMVQYEVK-WCLGNETRYHINKTVDGSTFSVVPFLVPGIRYSVEVAASTGAGPGV 862
Qy 854 YSKPISLFMDPTHVHPRAHPSGTHDGRHEGQDLTYHNGN-IPPGDINPTTHKKTIDY 912
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 863 KSEPFQIQLD-----SHGNPVSPEL-QVSLAQISDV 893

Qy 913 LSGP-----WLMVLVCIVLLVLVISAAISMVYFKRHQ----- 945
Db 894 VKQPAFIAGIGAACWILMVFSIWL-----YRHRKRNGLSSTYAGIRKVPST 942
Qy 946 ---MTKELGHLSSVSDNEITALNINSKESL-WIDHHRGW-RTADTKD-----SG 990
Db 943 FTPTVTYQRGGEAVSSGGPGLLNISEPATQPWLAD--TWPNTGNSHNDCSINCTASNG 1000
Qy 991 LSESKLLSHVNSSQ--SNYNNS-----DGGTDYAEVDTRNLTFYNCRKSPD--- 1035
Db 1001 NSDSNLTYSRPADCIANYNNQDNKQTNLMLPESTVYGDVLSNKNEMKTFNSPNLKD 1060
Qy 1036 -----NPTPYATTMII-----GTSSSE----- 1052
Db 1061 GRFVNPSPGQPTPYATTQLIQANLINNMNNGGDSSEKHKKPPGQKQEVAPIQYNIMEQN 1120
Qy 1053 -----TCTKTSISADKDSGTHSPYSADAFAGQVPAVPVVKSNLYQYPVEP----I 1098
Db 1121 KLNKDYRANDTILPTIPYHNSYDQNTGSSYNSSDRGSSSTSGSQGHKKGARTPKAPKQGGM 1180
Qy 1099 NWSEFLPPPEHPHPPSST-----YGAQGP-----ESSRKSSK 1132
Db 1181 NWADLPPPPAHPPPHSNSEYSMSVDESVDQEMPCVPVPPARMYLQDELEEEAERGPT 1240
Qy 1133 SAGSGISTNQSLNASIHSSSGGFSAGVSPQYAVA----CPPENVSNPLSAVAGGT 1187
Db 1241 PPVRGAASSPAAVSYS-HQST----ATLTPSQEELQMLQDCPED-----LGHMPHPP 1289
Qy 1188 QNRQITPTNQHPQLP-----AYFATTGP----- 1212
Db 1290 DRRRQ--PVSPPPRPISPHPTYGIISGLVSDMDTDAPEEEDEADMEVAKMQTRRL 1347
Qy 1213 -----GGAVPPNHL-----PFATQR 1227
Db 1348 LRGLEQTPASSVGDLESSVTGSMINGWSASEEDNISSGRSSVSSDGSFFTDADFAQAV 1407
Qy 1228 HAASEYQAGLNARCAQSRACNSDALATPSPMQPPPPVPEGWYQPVHPNSHPMHPTS 1287
Db 1408 AAAAEY-AGLKVARQMDAAGRHHFASQCP-RPTSPV-----STD 1447
Qy 1288 SNHQIYQCSSECDHRSQSSQSHKRLQLEEHGSSAKQRGHHRRA-----PVVQPC 1339
Db 1448 SN-----MSAAVIQKARPTKKQKHQ-----PGHLRREAYTDLPPPPVPPPA 1489
Qy 1340 MESENENMLAEYEQR 1354
Db 1490 IKSPSVQSKAQLEAR 1504

RESULT 8

QZ13

QZ13 PRELIMINARY; PRT; 1060 AA.
AC QZ13;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
DR EMBL; AF182037; AAF04558.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; lg; 5.
DR PRINTS; PR00014; FNTYPEIII.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Receptor.
FT NON_TER 1060 1060
SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C11E8542DA4 CRC64;

Query Match 20.2%; Score 1499; DB 11; Length 1060;
Best Local Similarity 34.7%; Pred. No. 1.7e-92;
Matches 361; Conservative 175; Mismatches 398; Indels 106; Gaps 28;

Qy 56 PRIIEHPTDLVVKNEPATLNCKVEGKPEPTI---EWFKDG-EPVSTNEKKSHRVQFKD 110
Db 31 PKXVEQSEVIVSKGKPNTPNWKQGRPFPTIGKVQRMVKPGWDTKDDSKVTQGCLLPS 90
Qy 111 GALFFYRTMQGKKEQ-DGGEYWCVAKNRVGQAVSRHASLQIAVLRRDDEPVEKPTRVAKG 169
Db 91 GSLFFLRIVHGRSRKPDDEGTYVCVARNYLGEAVSRNASLEVALRRDDEPVEKPTRVAKG 150
Qy 170 ETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNNLLISNVEPIDEG 229
Db 151 EPALLECQPPKGHPEPTIYKKDKRVIDE-----KEERISI-RGGKLMISNTRKSDAG 202
Qy 230 NYKICIAQLVGTRESSYAKLIVQVPMKEPKDQVMLYQGTATFHCSVGDDPPPKVLWK 289
Db 203 MYTCVGTNMVGERSDPAELTVFERPTFLRRPINQVLEDEPAEFRCVQGDPOPTVRWK 262
Qy 290 KEEGNIPVSRARILHDEKSLISNITPTDEGTYVCEAHNNVQGISARASLIVHAPPNFTK 349
Db 263 KDDADLPGRYDI-KDDYTLRIKKAISADEGTYVCIAENRVGKVEASATLTVRAPQFV 321
Qy 350 RPSNKKVGLNGVQLPCMASGNPPSVFVTEGKGVSTLMPFN---SSHGRQYVAADGTLQI 406
Db 322 RPRDQIVAGQRTVTPFCETKGNPQPAVFWKEGSGNLLFPNQPPQNSRCSVSPTGDLII 381
Qy 407 TDVRQDEGYVCSAFSVSDSTVRVFLQVSSV-DERPPPIIQIGPANQTLPGKSVATLP 465
Db 382 TNIQRSADAGYICQALTAVAGSILAKAQLVTDVLTDRPPPIIQGPINQTLAVDGTALLK 441
Qy 466 CRATGNPSRIKWFHDGHAVQAGNRSIIQG-SSLRVDDLQSDSGTYTCTASGERGETS 524
Db 442 CRATG-PLPVISWLKEGFTFLGRDPRTAIQDQGLTQIKNLRIISDTGTYTCVATSSGETS 500
Qy 525 WAATLTVEKPGSTSLHRAADPSTYPAPGTPKVLNVSRTSISLRWAKSDEKPGAVGPIIG 584
Db 501 WSAVLDTESGAT-ISKNYDTNLPGPSKQVTDVTKNSVTLSSWTG--TPGVL-PASA 556
Qy 585 YTYEYFSPDLQGTWIVAAHVRGDTQVTSISLTPGTSTYVFLVRAENTQGISVPSGLSNV 644
Db 557 YIIAEFSQSVSNWQTVANHVKTLLTYRGLRPNTIYLFMRVRAINPQGLSDPSPMSDPV 616
Qy 645 TIEADFAASANDLSAARTLLTGKSVELIDASAINASAVRLEWMLHVSADKEYVEGLRIH 704
Db 617 TQDIS-PPAQGVDRHQVQKELGVDVTRVLRNPNVVLPTTQVQVWT--VDRQPFQIQGYRM 673
Qy 705 YKDS--VPSAQYHSITVMDASAESFVGNLKKYTKYEFFLTPFFETIEGQPSNKALT 762
Db 674 YRQTSGLQASTVWQNLDAKVPTERSALVNLKGVTEIKVRPYNFEPQGMDSSEKTI 733
Qy 763 YEDVPSAPPDNI---QIGMNYQTAGVWRVTPPPSQHNGNLYGYKIEVSAGNTMKVLN 819
Db 734 TEEAPSAPPQSVTLTVGSHNSISVSWDPPADHNGIYQETKI-WCLGNETRFHINK 792
Qy 820 TLNATTSVLIENLTTGAVYSVRLNSFTKAGDGYPYSKISLFMDPTHVHPRAHPSGTH 879
Db 793 TVDATIRSVVIGLFPQIYRVVEAASSTAGVGKSESPQIII----- 835
Qy 880 DGRHEGQDLTYHNNGNIPPGDINPTTHKKTIDYLSGPWMLVLCIVLLVLVISAAISMVY 939
Db 836 GGRNE-VVITENNN-----SITEQITDVVKQPAFIAGIGGACWVILMGFSI-WLY 883

Qy 940 FKRK-----HQMTKELGHSVSDNEITA-LNINSKESLWIDHHRGWRTADTKDSSL 991
Db 884 WRRKKRGLSNYAVTFQRGDGLMSNGSRPLNTGDPNYWL-----ADSWP 931

Qy 992 SESKLLSHVNSQSNN-----SDGGDYAEVDTRNLTTFFNC 1030
Db 932 ATSLPVNNSNGPNEIGNFGRGDVLPVPGQGDKTATMLSDGAI-YSSIDFTTKTT-YNS 989

Qy 1031 RKSPDNPTPYATTMIIGTSS 1050
Db 990 SSQITQATPYATTQLHSNS 1009

RESULT 9
Q9VQ10

ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
T 01-MAY-2000 (TrEMBLrel. 13, Created)
T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG5481 PROTEIN.
GN CG5481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003586; AAF51373.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0031341; CG5481.

DR INTERPRO; IPR001412; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 1.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEI11.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
SQ SEQUENCE 823 AA; 89715 MW; 36FC0B91F36F2F19 CRC64;

Query Match 19.3%; Score 1430; DB 5; Length 823;
Best Local Similarity 36.4%; Pred. No. 5.2e-88;
Matches 301; Conservative 141; Mismatches 249; Indels 136; Gaps 19;

Qy 64 DLVVKKNEPATLNCVEGKPEPTIEWFKDGEVSTNEKSHRVQFKDGLFFYRTMQGKK 123
Db 2 DTTVPKNDPTFNCAEGNPTPTIQWFKDRELKT-DTGSRLMLPAGGLFPLKVIHSRR 60
Qy 124 EQDGGYWCYAKNRVGOAVSRHASLQIAVLRRDPRVEPKDTRVAKGTALLEGCPKGP 183
Db 61 ESDAGTYWCEAKNEFGVARSRNATLQVAFRLDRLEPANTRVAQGEVALMECAGPRGSP 120
Qy 184 EPTLIWIKDGVPLDDLKAMSGASSRVRIVDGNLLISNVEPIDEGNYKCIAQLVGTRE 243
Db 121 EPQISWRKNG-----QTLNLVGNKRIRIVDGNLAIQEARQSDGGRYQCVKVVVGTRE 174
Qy 244 SSYAKLVQVKPYFMKEPKDQVMLYQGTATFHCSSVGDPKPKVLWKK--EEGNIPVSRAR 301
Db 175 SATAFLLVHVRPFLIRGPQNTAVGVSSVVFQCRIGDPLDVLWRRATSGGNMP-RRVH 233
Qy 302 ILHDEKSLISNITPTDEGTYVCEAHNNVQISARASLIVHAPPNFKRPSNKKVGLNGV 361
Db 234 VLED-RSLKLDVTLDEMGYTCADNAGVGITATGILTVHAPPKVRPNQLVEIGDE 292
Qy 362 VQLPCMASGNPPPSVFWTKEGVSTLMFPNSSHGRQYVA----ADGTLQITDVRQDEGY 417
Db 293 VLFECQANGHPRTLYWSVEGNSLLLPYGRDGRMEVTLTPEGRSVLSIARFAREDSGV 352
Qy 418 V-CSAFSVVDSSTVRVFLQVSSVDERPPPIIQIGPANQLPKGSVATLPCRATGNPSPRI 476
Db 353 VTCNALNAVGSVSRVSVDTQFELPPPIIEQGPVNTLPVKSIVVLPCLRTLTPTVPQV 412
Qy 477 KWFHDGHA--VQAGNRYIIQSSSLRVDDLQ-LSDSGTYTCTASGERGETSWAATLTVK 533
Db 413 SWYLDGIPIDVQEHERRNLSAGALTISDLQRHDEGLTYCVASNRNGKSSWSGYLRD 472
Qy 534 PGSTSL--HRAADPSTYPAPPPTKVLNVSRISLWAKSSEKPGAVGPIIGTYVEYFS 591
Db 473 PTNPNIKFFRAPELSTYPGPKQMVKEGKSVTLNWKVGS--SLVGYVIEMFG 530
Qy 592 PDLQGWIVAAHRVGDQVITISGLTPGTSYVFLVRAENTQGISVPSGLSN--VIRKTEAD 649
Db 531 KNETDGVAVGTRVQNTTFTQTGLLPVNYFLIRAEHSHGLSLPSPMSEPTVGTYSSE 590
Qy 650 -----FDAA-----SANDLSAAR-TLLTGKSVELIDASAINASAVRLE 686
Db 591 NHEFTLMPFSLIHFDLPHQRYFNSGLDLSEARASLLSGDVVLSNASVVDSTSMKLT 650
Qy 687 WML-----HVSA-----DEKYVEGLRIHYKADSP----- 711
Db 651 WQVCNRLTDGSIAPHSIAHRLIRASFLMQIINGKYVEGYVYARQLPNVNNPAPV 710
Qy 712 -----SAQYHSIT 719
Db 711 TSNTNPLLGSTTSASASASASALISTKPNIAAACKRDGETNQSGGAPPLNTKYRMLT 770
Qy 720 VMD-ASAEFVGNLKKYKYEFFLTPFFETIEGQPSNKTALTYED 765
Db 771 ILNKGASSCITGLVQYTLIEFFIVPYKSVGKPSNRIARTLED 817

RESULT 10
Q92214


```

ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN ZK377.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,

```

```

RESULT 12
001632
ID 001632      PRELIMINARY;      PRT; 874 AA.
AC 001632;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA CEESC12R.

```

```

Qy 442 RPPPIIQIGPANQTLPGKGSVATLPCRATGNPSPRIKWFHDGHAQV-AGNRYSIIQSSSLR 500
      :||| :| :||| || |||::|| :| :| :| :| :|
Db 27 KPPPTIEHGHQNTLMVGSSAILPQASGKPTPGISWLRDGLPIDITDSRISQHSTGSLH 86
      :||| :| :||| || |||::|| :| :| :| :| :|
Qy 501 VDDLQLSDSGTYTCTASGERGETSAAATLTVEKPGSTS-LHRAADPSTYPAPPGTKPVLN 559
      :||| :| :||| || |||::|| :| :| :| :| :|
Db 87 IADLKRPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPTQPIIYN 146
      :||| :| :||| || |||::|| :| :| :| :| :|
Qy 560 VSRSTSISLRWAKSQEKPQAVGPIIGYTVFEYSPDLQTGWIVAAHRVGDQTVTISGLTPGT 619
      :| :| :| :| :| ||| || :||| :| :| :| :| :|
Db 147 VDTIEVELHW--NAPSTSGAGPTGYIIQYISPDLGQTFWNIPDYVASTEYRIKGLKPSH 204
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 620 SYVFLVRAENTQGISVPSGLSNVKTIEADFDAASAN---DLSAARTLLTGKS-VELID 674
      ||::||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 205 SYMFVIRAENKIGTGPSPVSALLVTISKPAQVALSDKNKMDMAIEKRRLTSEQLKLEE 264
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 675 ASAINASAVRLWMLHVSADKEYEGLRIHYK-DASVPSAQYHSITVMDASAESFVVGNL 733
      ||::||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 265 VKTINSTAVRLFWKKRKL--EELIDGYIKWRGPRPTNDNQY--VNVTPSPSTENYVVSNL 320
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 734 KKYTKYEFFLTFF---PETIEGQPSNSKTALTYEDVPSAPPDNIQIGMYNQTAGVWRWTP 790

```

Qy 157 FRVEPKDTRVAKGETALLEGCPKPGIPEPTLIWIKDGVPLDDLKAMSGPAGSSRVRIVDGG 216
| | | | : : : : : : : : : : : : : : : : : :
Db 24 FLVFPVDTLSVRGSSVILNCS-ATSESPSNI EWKKDGT-----FLNLVSDDRQLLPDG 76

Qy 217 NLLISNV-----EPIDEGNYKCIA--QNLVGTRRESSYAKLIVQVKPYFMKEPKDQMVLV 268
: | | | : | | | : : | | | | | | : : :
Db 77 SLFISNVVSHKHNP-DEGFYQC VATVDNL-GTIVSR TAKLAVAGLP RFTSQPEPS SIV 134

Qy 269 GQTATFHCSVGGDP PPKVLWKKEEKNIPVSRARILHDEK-----SLEISNITPTDEGT 321
: : : | : | | : : : : | : : : : | | | | | |
Db 135 GNSGILNCEVINADLVPPVRWEQ-----NRQPLL LLDRI VKLPSTGLVISNATEGDEGL 187

Qy 322 YVC-----EAHNNGVQISARASLIVHAPPNF TKRPSN--KKVGLNGVVQLPCM 367
| | : | | | : | : | : | : | : | : | : | :
Db 188 YRCIVESGGPPKFSDEAEKVQESEM L DV-----FLMRPSSMIKVIGQSAV--LPCV 240

Qy 368 ASGNPPPSVFVTIKEGYSLTMFPNSSHGROYAADGT LQITDVREGDEGYVCSAFSVVDS 427
| | | | : : : | | : | : | : | : | : | : | :
Db 241 ASGLPAPVIRWMK---NEDVLT ESSGR LALLAGGSLEISDVTEDDAGTYFC---VADN 293

Qy 428 STVRVLQVSSVDERPPP I IQIGPANQTL PKGSVATLPCRATGNPSPRIKWDFGHAVQA 487
: | : : : | : : | : : | : : | : | : | : | : | :
Db 294 GNKTIEAQ AELTVQVP EFLK-QPANIYARESM DIVFECEVTGKPAPT V KWKNKG DVIPI 352

Qy 488 GNRYSIIOGSSLRVDDLQSLDSGTYTCTASGERGETSWAATLTVEKPGSTSLHRAADPST 547
: : : : : | : | | | : : : | : : : : : : : :
Db 353 SDYFKIVKEHLQVLGLVKSDEGYQCI AENDVNAQAQA QLIILE-----HAPATTGP 405

Qy 548 YPAPPGTPKVL NVSR T SISLRWAKS QEKPGAVGIIGTYTEYFSPDLQTGWIVAHRVGD 607
| : | : | : | : | : | : | : | : | : | : | : | :
Db 407 LP SAPRDVVASLVSTRFIKLWRTPSDPH--GDNL TYSVFYTK EGVARERVENTSQPGE 464

Qy 608 TQVTISGLTPGTSYFVLVRAENTQG ISVPSGLSNVIKTIEADFDAASANDLSAARTLLTG 667
| | | | : | : | : | : | : | : : | : : : | :
Db 465 MQVTIQNLMPATVYIFKVMAQNKHG---SGESSAPLRVE-----TQ 502

Qy 668 KSVEL-----IDASAINASAVRL EWM LHVSAD EKYVEGLRIHKDASVP SAQYHSITVM 721
: | :
Db 503 PEVQLPGPAPNIRAYATSPTSITVWETPLSGNGE-IQNKLYMEKGTDKQE----DV 556

Qy 722 DASAESFVGNLK KYEFFLTPFFETIEGQPSNSKALTIEDVPSAPPDNIQIGMYNQ 781
| : : : : | | | : : : : : : : : : : : : : :
Db 557 DVSSHYSYTINGLKKYTEYFRVVAYNKHGPGVSTQDVAVRTSDVPSAAPQNLSEVRNS 616

Qy 782 TAGWVRWTPPPSQHHNGLYK IEVSAGNTMKVLANMTLNATTT SVLLNNLTGAVYSV 841
: : : | | : | : | : | : | : : : : : : : : :
Db 617 KSIIVHWQPSSATQNGQITGYKIRYKASRSKSDVTVTVTGTQLSQLGLEGLDRGTEYNF 676

Qy 842 RLNSFTKAGDGPYSKPIS--LFMDPTHVHPRAHPSGTHDGRHEGQDLYTHNNGNIPP 899
: : : | : | : : : : : : : : : : : : : : : :
Db 677 RVAALT VNGTGPATDWLSAETPESDLD ES RVPE-PSSLH-----715

Qy 900 DINPTTHKKT DYLSGPWLMLVCIVLLVLVISAAIMVYFKRKHQMTK--ELGHL SVVS 957
: | : | : | : | : | : | : | : | : | : : : :
Db 716 -VRP-----LVT SI VSWTPPENQINIVRGYAIGY----744

Qy 958 DNEITALNINSKESLWIDHRRGWRTADTDKDSLSESKLLSHVNSSQSNNYNSDGG----1013
:
Db 745 -----GIGSPAHTIKVDYKQRYTTIENLDPS-----SHYVITLKAFNNVWGEGIPLY 791

Qy 1014 -----TDYAEVDTRNLTFYNCRKSPDNPTYATTMI--IGTSSSETCKTTSIS-A 1062
| : | : | : : : | : | : : | : : : | : | :
Db 792 ESAVTRPHDI SEVDFVI-----NAPYTYPVDPPTMMPPGVGQASILSHDTRITWA 844

Qy 1063 DKDSGTHSPYSDA-----FAGQVP AVPVVKS-----NYLQYVPEPINWSEF-----1103
: | : | : : : : : : : : : : : : : : : : :
Db 845 DNSLPKHOKI IDSRYYTVRWKTNIPANTKYKANANATLSYLVTGLKPNLYFEESVMVTGK 904

SQ SEQUENCE 1377 AA; 150637 MW; E514ED8ABD1A63A9 CRC64;

QY 124 EQDGGEYMCVAKNR-----VGQAVSRHASLQIAVLRDD----- 156
 |::| |::| |::| |::| |::|
 Db 4 EREAGRLCLTSSSRRCPPPLPLLLPLLLGRPASGAATKSGPRQSQGASVRTFTF 63
 QY 157 --FRVPEKDTRVAKGETALCEGPKPGKIEPTLIWIKDGVPLDDLKAMSGFASSRVRIVD 214
 |::| |::| |::| |::| |::|

[illegible]

```

Db      280  IKWMKNEBALDT-----ESSERLVLLAGSLEISDVTEDDAGTYFC-----IADNGNETIE 330
Qy      434  LQVSSVDERPPPIIQIGPANQTLPGKSVATLPCRATGNPSPIKWFHDGHAVQAGNRYSI 493
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      331  AQAEITVQAQPEFLK-QPTNIYAHESMDIVFECEVTGKPTPTVKWKNGDMVIPSDYFKI 389
Qy      494  IQGSSSLRVDDLQLSDSGTYTCTASGERGETSAAWALTVEKPGSTS LHRAADPSTYPAPPG 553
      : : : : : : : : : : : : : : : : : : : : :
Db      390  VKEHNQLVGLGLVKSDEGYPQICIAENDVGNAQAGALITILE-----HAPATTGPLPSAPR 443
Qy      554  TPKVLNVSRTSISLRWAKSQEKPAGVGPIIGITYVEYFSPDLQTCWIVAHRVGTDTQVTIS 613
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      444  DVVASLSTRFIKLTWRTPASDPH--GDNLTYSVFYTKEGIARERVENTSHPGEMQVTIQ 501
Qy      614  GLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIADFDAASANDLSAARTLLTGKSVEL- 672
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      502  NLPMPATVIFRVMQAKHG---SGESSAPLRVE-----TQPEVQLP 539
Qy      673  ----IDASAINASAVRLEWMLHVSADKEYVEGLRIHYKASVPSAQYHSITVMDASAES 727
      : : : : : : : : : : : : : : : : : : : : :
Db      540  GPAPNLRAYAASPTSITVTWETPVSGNGE-IQNYKLYMEKGTDKEQ-----DVDVSSH 593
Qy      728  FVVGNNKKYTKYEFFLTFFPETIEGQPSNKSALTIEDVPSAPPDNIQIMYNTAGWVR 787
      : : : : : : : : : : : : : : : : : : : : :
Db      594  YTINGLKKYTEYSFRVVAYNKHGPGVSTPDVAVRTLSDVPASAPQLNSLEVRNKSIMIH 653
Qy      788  WTPPPSQHHNGNLGYKIEVSAGNTMKVLNMTLNATTSTVLLNNLTGAVYSVRLNSPT 847
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      654  WQPPAPATQNGQITGYKIRYKASRKSVDVETLVSQGLSGLIEGLDRGTEYNFRVAALT 713
Qy      848  KAGDGPYSKPS--LFMDPTHVHPPRAHPSGTHDRGREGQDLTYHNNGNIPPGDINPTT 905
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      714  INGTGPAIDWLSAETFESDLDETRVEV-PSSLH-----VRP-- 749
Qy      906  HKKTTDYLSGFWMLVLCIVLLVLVISAAMVYFKRKHQMTK--ELGLHSVVDNEITA 963
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      750  -----LVTISVVSPTPENQNIIVRGYAIQY-----G 776
Qy      964  LNINSKESLWIDHHRGWRTADTDKDSGLSESKLLSHVNSSSQSNYNNSDGG----- 1013
      : : : : : : : : : : : : : : : : : : : : :
Db      777  IGSPhAQTKVDYKQRYTIEIENLDP-----SHYVITLKAFNNVGEGLPIYESAVTR 828
Qy      1014  --TDYAEVDTRLNTTFYNCRKSPDNPTPYATTMI--IGTSSSETCKTKTTSIS-ADKDSGT 1063
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      829  PHTDTSEVDLFI-----NAPYTPVPDPTPMMPVGVQASILSHDTIRITWADNSLPK 881
Qy      1069  HSPYSDA-----FAGQVPAPVPVKS-----NYLQYPVEPINWSEF----- 1103
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      882  HQKITDSRYITVRKNTINIPANTKYKANATLSYLVLTGKPNLTLYEFSVMVTKGRRSSTW 941
Qy      1104  -----LPPPEHPPPSSTYGYAQGSPE-----SRKSKKSAGSGI--STNQSIL 1143
      : : : : : | : : : : : | : : : : : | : : : : :
Db      942  SMTAGCTTFELVPTSPPKDVTVVSKEGPKTIIVNQWPPSEANGKITGYIITYSTD--V 998
Qy      1146  NASIH-----SSSSGGFSAWGVSPQYAV-ACPPEN 1174
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      999  NAEIHDWVIEBPVGNRLTHQIQELTLDTPYYPKIQARNKSGMGPMSEAVQFRTPKADSSD 1051
Qy      1175  VYSNPLSAVAGTQNRQITPTNQHPQLPAYFATTGPGGAVPPNHL----- 1221
      : : : : : | : : : : : | : : : : : | : : : : :
Db      1059  KMPNDQASGSGGKGRSLPDLGSDYKPPMSGSNSPHGSPSTPLDSNMLLVIIVSVGVITIV 1111
Qy      1222  -----PPATQRHAASEYQAGLNAARCAQSRACNSCDA---LATPSMPQPP----- 1261
      | : : : : : | : : : : : | : : : : : | : : : : :
Db      1119  VVVIIVAFCTFRRTTSHQKK-----KRAACKSVNGSHKYKGNKDKVPPDLWIHHER 1161
Qy      1264  -----PPVPVEGWYQPVHPNHPMPEHTSSNHQIYQCSSECDHSSRSSQSHKRLQL 1311
      : : : : : | : : : : : | : : : : : | : : : : :
Db      1170  LELKPIDKSPDPNPIMTDPITPRNSQDITPV-----DNSMSDNIHQRRNSY 1211
Qy      1316  EEHGSS-----AKQRGGHRRRAPV----VQPCMESENENMLAEYEQRYTSDCCNNSR 1361
      | : : : : : | : : : : : | : : : : : | : : : : :

```

```

ID 000340          PRELIMINARY;          PRT; 1461 AA.
C 000340;
T 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE NEOGENIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eute
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Hom
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97312699; PubMed=9169140;
RA Vielmetter J., Cheng X.N., Miskevich F., Lane R.P., Yama
RT Korenberg J.R., Dreyer W.J.;
RT "Molecular characterization of human neogenin, a DCC-rel
RT and the mapping of its gene (NEO1) to chromosomal positi
RL q23.";
RL Genomics 41:414-421(1997).
DR EMBL; U72391; AAC51287.1; -.
DR HSSP; P02751; ITTF.
DR INTERPRO; IPR000531; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII'.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
SQ SEQUENCE 1461 AA; 160015 MW; 4AADF1EEBCAFD82C CRC64;

```

Query Match 9.6%; Score 712.5; DB 4; Length 1461;
Best Local Similarity 22.3%; Pred. No. 2.8e-39;
Matches 313; Conservative 184; Mismatches 564; Indels 341; Gaps 48;

```

Qy      157 FRVEPKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSGASSRVRIVDGG 216
      | | | | | : | : | | | | : | : | : | : |
Db      55 FLVEPVDTLISVRGSSVILNCS-AYSEPSKIEWKDKGT-----FLNLVSDDRQLLPDG 107

Qy      217 NLLISNV-----EPIDEGNYKICIAQ-NLVGTRESSYAKLIVQKPYFMKEPKDQVMYLG 269
      : | | | | : | | | | : | : | | | | | | : | : |
Db      108 SLFISNVVHSHKNPK-DEGYQCVATVESLGTIISRTAKLIVAGLPRPTSQPESSVYAG 166

Qy      270 QTATFHC SVGGDPPPKVLWKKEEGNIPVSRARILHDEK-----SLEISNITPTDEGTY 322
      | : | | | | | : : : | | : : | | | | | | |
Db      167 NNAILNCEVNA DLVPFVRWEQ-----NRQPLLLDDRVIKLP SGM LVI SNAT EGDGGLY 219

Qy      323 VCEAHN-NVGQISARASLI VHAPPN-----FTKRPSKNKVLGNLVQLPCMASGNPPPS 375
      | : : : | | | | | | : | : | : | | | | |
Db      220 RCVVESGGPKYSDEVELKVLDPDEVISDLVFLKQPSPLVRVIGQDVLPCVAGSLPTPT 279

Qy      376 VFWTK--EGVSTLMPNNSHGRQYVADGTLQITDVROEDEGYVCSAFSVSDSSTVRVF 433

```

Db 1216 RGHESEDSMSTLAGRRGMRPKMMMPFDSQPPQPVISAHPIHSLDNPHHHFHSSSLASPAR 1275

Qy 1366 EGDTCSCSEGSCLYAEAGEPAP 1387

Db 1276 -----SHLY-HPGSPWP 1286

Search completed: January 22, 2001, 12:51:13
Job time: 1874 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:17:03 ; Search time 233.01 Seconds
(without alignments)
202.659 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTVPKNDP.....RSLLSNSGSGTSSQAGHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7272	100.0	1381	20	Y13564	Drosophila Robo 2
2	7256.5	99.8	1380	20	Y08402	Drosophila sp. ROB
3	1786.5	24.6	1395	20	Y13563	Drosophila Robo 1
4	1786.5	24.6	1395	20	Y08401	Drosophila sp. ROB
5	1498	20.6	1651	20	Y13566	Human Robo 1 polyp
6	1491	20.5	1649	20	Y08404	Human ROBO1 protei
7	1344.5	18.5	1297	20	Y13565	C. elegans Robo po
8	1344.5	18.5	1297	20	Y08403	C. elegans ROBO pr
9	1261	17.3	753	20	W83927	Human T85 protein.
10	557.5	7.7	1728	12	R13144	Deleted in Colorec
11	551.5	7.6	1447	16	R68553	Deleted in colorec
12	551.5	7.6	1447	20	Y33498	Human DCC protein.

13	548.5	7.5	1192	19	W57900	Protein of clone C
14	543.5	7.5	1028	19	W29667	Homo sapiens DL165
15	538	7.4	1571	19	W42087	Human Down syndrom
16	538	7.4	1910	19	W42086	Human Down syndrom
17	536.5	7.4	1018	15	R63759	Human contactin (E
18	536.5	7.4	1018	17	R87028	Human contactin.
19	531.5	7.3	1496	20	W81030	Melanoma associate
20	531.5	7.3	1496	21	Y70469	Human p53 target m
21	524	7.2	1018	18	W06485	Rat contactin liga
22	509.5	7.0	1299	21	Y40439	Human Nr-CAM prote
23	509	7.0	1257	20	W74152	Human L1 cell adhe
24	495.5	6.8	1897	21	Y81785	Human protein tyro
25	495.5	6.8	1897	21	Y56100	LAR tyrosine phosph
26	495	6.8	1304	19	W59994	Human neural cell
27	493.5	6.8	1911	16	R71726	Human PTP-OB. Hom
28	493.5	6.8	1911	18	W27225	Human protein tyro
29	493.5	6.8	1911	20	W94027	Human protein tyro
30	484	6.7	3117	21	Y53667	Sequence gi/332818
31	475	6.5	1242	19	W52287	Rattus norvegicus
32	473.5	6.5	434	20	Y13567	Human Robo 2 polyp
33	473.5	6.5	434	20	Y08405	Human partial ROBO
34	448.5	6.2	4412	21	Y53666	Sequence gi/101742
35	438	6.0	1225	19	W52289	Homo sapiens cdo t
36	431	5.9	1501	16	R72858	Rat receptor type-
37	420.5	5.8	761	17	R92255	Neural cell adhesi
38	415.5	5.7	1251	19	W37778	Rattus norvegicus
39	412.5	5.7	1070	18	W08747	Human colon carcin
40	388.5	5.3	848	21	Y88565	Human NCAM 140kd i
41	382	5.3	1853	21	Y53668	Protein 608 sequen
42	382	5.3	2387	21	Y53665	Mechanical stress
43	382	5.3	2597	21	Y53664	Mechanical stress
44	370.5	5.1	1125	19	W52288	Rattus norvegicus
45	370.5	5.1	1139	19	W37779	Rattus norvegicus

ALIGNMENTS

RESULT 1
Y13564
ID Y13564 standard; Protein; 1381 AA.
XX
AC Y13564;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W0925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338003/28.
DR N-PSDB; X55768.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 34-38; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 1381 AA;

Query Match 100.0%; Score 7272; DB 20; Length 1381;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GENPRIIEHPMDTTPKNDPFTFNCQAEQNPTPTIQWFKDRELKTDGSHRIMLPAGGL 60
Db 1 genpriiehpmdttvpkndpftfncqaeqnpptptiqwfkdgrelktdtgsrhlmpaggl 60

Qy 61 FFLKVIHSRRESDACTYWEAKNEFGVARSRNATLQVAVLRDEFRLPANTRVAGGEVAL 120
Db 61 fflkvihsrresdagtyweaknefgvarsrnatlqvavlrdefrlpantrvaggeval 120

Qy 121 MECGAPRGSPEPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDGGRYQCVKNV 180
Db 121 mecgaprgspegqiswrkngqtlnlvgnkririvdggnlaiqearqsddgryqcvknv 180

Qy 181 GTRESATAFLKVHVRPFLIRGPQNTAVVSSVVFQCRIGGDPLPDVLRRTASGGNMPL 240
Db 181 gtresataflkvhrpflirgpgnqtavvssvfvqcriggdplpdvlwrrtasggnmpl 240

Qy 241 RKFSLWHSASGRVHVLDRSLKLDVLTEDMGEYTCEADNAVGGITATGILTVHAPPKFV 300
Db 241 rkfswlhasgrvhvledrslklldvltedmgeytceadnavggitatgiltvhappkf 300

Qy 301 IRPKNLVEIGDEVLFECQANGHPRTIYWSVEGNSLLPGYRDGRMEVTLTPEGRSVL 360
Db 301 irpknqlveigdevlfecqanghrptiylwsvegnslilpgyrdgrmevtiltpegrsvl 360

Qy 361 SIARFAREDSGKVVTNCALNAVGSVSSRTVSVDTQFELPPPIIEQGPVNQTLPVKSIVV 420
Db 361 siarfaredsgkvvtncalnavgsvssrtvsvdtqfelpppieqgpvnqtlpvksivv 420

Qy 421 LPCRTLGTTPVPQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDEGLYTCVASNRN 480
Db 421 lpctrlgttpvpqvsyldgipidvqeherrnlsdagaltisdqrhedeglytcvasnrn 480

Qy 481 GKSSWSGYLRDTPNPNKFFRAPELSTYGPQPGKPMVEKGENSEVTLSTWRNSKVGGS 540
Db 481 gksswsyglrdtptnnpnkffrapelstygppgkpmvekgensvltswrnskvgs 540

Qy 541 SLVGYVIEMFGKNETDGVAVGTRVQNTTFTQTGLPGVNYFFLIRAENSHGLSLSPMS 600
Db 541 slvgviemfgknetdgvavgtrvqnttftqtgllpgvnyffliraenshglslspms 600

Qy 601 EPITVGTTRYFNSGLDSEARASLLSGDVVLSNASVVDSTSMKLTWQIINGKYVEGFYV 660
Db 601 epitvgttryfnsldsearasllsgdvvlsnasvvdstsmkltwqiingkyvegfyv 660

Qy 661 ARQLPNPNNPAPVTSNTNPLLGSTSTASASASALISTKPNIAAGKRDGETNQSG 720
Db 661 arqlpnpnnpapvtsntnpllgststasasasalistkpniaagkrdgetnqsg 720

Qy 721 GGAPTPLNTKYRMLTILNGGASSCTITGLVQYTLTYEFFIVPFYKSVGKPSNSRIARTL 780
Db 721 ggaptplntkyrmltilnggassctitglvqytltyeffivpfykgpsnsriartl 780

Qy 781 EDVPSEAPYGMEALLNNSAVFLKWKAPELKDRHGVLLNYHIVRGIDTAHNSRILTNV 840
Db 781 edvpseapygmeallnnsavflkwapelkdrhgvllnyhivrgidtahnsrilt 840

Qy 841 TIDAASPTLVIANLTEGVMYTVGVAAAGNAGVGPYCPATLRLDPITKRLDPFINQRDHV 900
Db 841 tidaasptlvianltegvmtyvgvaagnagvgpycpatlrlldpitrldpfinqrhdv 900

Qy 901 NDVLTQPWFIIILGAILAVLMLSGAMVFKRKHMMKQKQKQKQKQKQKQKQKQKQKQKQKQ 960
Db 901 ndvltqpwfiillgailavlmlsgamfvkrkhmmkqkqkqkqkqkqkqkqkqkqkqkqkq 960

Qy 961 ARNGNGYWLDSSTGGMVWRPSPGGDSLEMQKDHADIAPVCGAPGSPAGGTSSTGGSGGA 1020
Db 961 arngngywlssstggmvwrpspggdslemqkdhadiapvcgagpsaggtssggsgga 1020

Qy 1021 GSGASGDDIHGGHGSERNQRYVGEYSNIPTDYAEVSSFCAPSEYGRHGNASPAPYAT 1080
Db 1021 gsgasgddihgghgsernqryvgeysniptdyaevsfkapseygrhgnaspapyat 1080

Qy 1081 SSILSPHQQQQQQPRYQQRVPVGYGLRPMHPHYQQQQHQQQQQAQTHQHQALQQHQ 1140
Db 1081 ssilspHQQQQQQPRYQQRVPVGYGLRPMHPHYQQQQHQQQQQAQTHQHQALQQHQ 1140

Qy 1141 LPPSNLYQMQSTTSEIYPTNTGPSRSVSEYQYYPKDKQRHITHITENKLSNCHTYEAPG 1200
Db 1141 lppsnlyqmqsttseiypntgpsrsvyseyyypkdkqrhihitenklsnchtyeapg 1200

Qy 1201 AKQSSPISQFASVRRQQLPPNCISGRESARFKVLNTDQKQNNQLLDLGGSSMCYNGLA 1260
Db 1201 akqsspissqfasvrrqqlppncisgresarfkvlntdqkqnnqlldlgssmcynla 1260

Qy 1261 DSGCGGSPSPMAMLSHEDHALYHTADGLDDMERLYVKVDEQQPPQQQQQLIPLVPQH 1320
Db 1261 dsgcggspspmamlshedahalyhtadglddmerlykvvdeqqppqqqqqliplvpqh 1320

Qy 1321 PAEGLHLSWRNQSTRSSRNQCECIKEPSELAPGSVASERSLLSNSGGTSSQPAGHN 1380
Db 1321 paeghlswrnqstrssrnqcecikepseliapgsvasersllsnsggtssqpagn 1380

Qy 1381 v 1381
Db 1381 v 1381

RESULT 2

Y08402

ID Y08402 standard; Protein; 1380 AA.

XX

AC Y08402;

XX

DT 24-JUL-1999 (first entry)

XX

DE Drosophila sp. ROBO2 extracellular domain protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
cell morphology; screening assay.

XX

OS Drosophila sp.

XX

PN W09920764-A1.

XX

PD 29-APR-1999.

XX

PF 20-OCT-1998; '98WO-US22164.

XX

PR 14-NOV-1997; 97US-0971172.

PR 20-OCT-1997; 97US-0062921.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman CS, Kidd T, Mitchell KJ, Tear G;

XX

DR WPI; 1999-312615/26.

DR N-PSDB; X57251.

XX

PT Robo polypeptides, a new immunoglobulin superfamily member

XX
PS Claim 1; Page 52-56; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1380 AA;

Query Match 99.8%; Score 7256.5; DB 20; Length 1380;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1380; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 GENPRIIEHPMDTTPKNDPFTFNCQAEGNPTPTIQWFKDGRLEKTDGSHRIMLPAGGL 60
Db 1 genpriiehpmdttvpkndpftfncqaegnptptiqwfkdgrelktdtgshrimlpaggl 60
Qy 61 FFLKVIHSRRSDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFRLPANTRVAQGEVAL 120
Db 61 fflkvihsrrsdagtywceaknefgvarsrnatlqvavlrdefrlpantrvaggval 120
Qy 121 MECGAPRGSPEPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDGGRYQCVKVVV 180
Db 121 mecgaprgspegpqi swrkngqtlnlvgnkririvdggnlaiqearqsdggrycvkvvv 180
Qy 181 GTRESATAFLKVHVRPFLIRGPQNTAVVGSSVVFQCRIGGDLPLDVLWRRTASGGNMPL 240
Db 181 gtresataflkvhvrpflirgpnqtavvgssv v f q c r i g g d l p d v l w r r t a s g g n m p l 240
Qy 241 RKFSWLHSASGRVHVLEDRSLKLDVLTLEDMEGYTCEADNAVGGITATGILTVHAPPKFV 300
Db 241 rkfswlhasasgrvhvledrslkldvltledmgytceadnavggitatgiltvhappkfiv 300
Qy 301 IRPKNLVEIGDEVLFECQANGHPRTLYWSVEGNSLLLPYRDRGMEVTLTPEGRSVL 360
Db 301 irpknqlveigdevlfecqanghprptlywsvegnsslllp y r d r g m e v t l t p e g r s v l 360
Qy 361 SIARFAREDSGKVVTCTNALNAVGSVSSRTVSVDTQFELPPPIIEQGPVNQTLPVKSIVV 420
Db 361 siarfaredsgkvvtctnalnavgs v s s r t v s v d t q f e l p p p i e q g p v n q t l p v k s i v v 420
Qy 421 LPCRTLGTTPVPQVSWYLDGIPIDVQEHERRNLS DAGALTISDLQRHEDGLYTCVASNRN 480
Db 421 lpcrtlgttpvpqvs w y l d g i p i d v q e h e r r n l s d a g a l t i s d l q r h e d g l y t c v a s n r n 480
Qy 481 GKSSWSGYLRDLTPNPNKFFRAPELSTYPPGPKPMVEKGNSVTLSTWTRSNKVGG 540
Db 481 gksswsgylrldtptn p n k f f r a p e l s t y p p g p k p m v e k g n s v t l s t w t r s n k v g g s 540
Qy 541 SLVGVIEMFGKNETDGWVAVGTRVQNTTFTQTGLLPGVNYFFLIRAENSHGLSLPMS 600
Db 541 slvgviemfgknetdgwvavgtrvqnttftqtgl l p g v n y f f l i r a e n s h g l s l p m s 600
Qy 601 EPITVGTTRYFNSGLDSEARASLLSGDVVELSNASVVDSTSMKLTWQINGKYVEGFYVY 660
Db 601 epitvgttryfns g l d s e a r a s l l s g d v v e l s n a s v v d s t s m k l t w q i n g k y v e g f y v y 660
Qy 661 ARQLPNPVI VNNPAPVTSNTNPLLGSTSTASASASASALISTKPNIAAGKRDGETNQSG 720
Db 661 arqlpnpi v n n p a p v t s n t n p l l g s t s t a s a s a s a l i s t k p n i a a g k r d g e t n q s g 720
Qy 721 GGAPTPLNTKYRMLTILNGGASSCTITGLVQYTYLIEFFIVPFYKSVGKPSNSRIARTL 780
Db 721 ggaptplntkyrmltilnggassctitglvqy t y l i e f f i v p f y k s v e g k p s n s r i a r t l 780
Qy 781 EDVPSEAPYGMERALLNSSAVFLKWAPELKDRHGVLNHYHIVRGIDTAHNSRILTNV 840
Db 781 edvpseapygmeallnssavflk w a p e l k d r h g v l n y h i v r g i d t a h n s r i l t n v 840

Qy 841 TIDAASPTLVLANLATEGVMTYVGAAGNAGVGPYCPATLRLDPITKRLDPFINQRDHF 900
Db 841 tidaasptlvlanlategvmytvgvaagnagvgpycpatrlrldpitrldpfinqrhdv 900
Qy 901 NDVLTQPWFIIILGAILAVLMLSGAMVFKRKHMMSKQALNTRMGNTSDVLKMPSL 960
Db 901 ndvltqpwfiillgailavlmlsfgamvfk r k h m m k s a l n t m r g n t s d v l k m p s l s 960
Qy 961 ARNGNGYWLDSSTGGMVWRPSPGGDSLEMKDHIADYAPVCGAPSPAGGTS SSGSGGA 1020
Db 961 arngngywl dsstggmvwrpspggdslemkdhiadyapvcgappsgaggts ssgsgga 1020
Qy 1021 GSGAGGDDIHGGHGSERNQRYVGEYSNIPTDYAEVSSFGKAPSEYGRHGNASPAPYAT 1080
Db 1021 gsgagggddihggghgsernqrryvgeysniptdyaevssfgkapseygrhgnaspapyat 1080
Qy 1081 SSILSPHQQQQQQPRYQQRVPVGYGLQRPMPHYQQQQHQQQQAQTHQQHQAQQHQ 1140
Db 1081 ssilsp h q q q q q p r y q q r v p v g y g l r p m p h y q q q q h q q q a q t h q q h a l q q h q 1140
Qy 1141 LPPSNIYQQMSTTSEIYPTNTGPSRSVYSEYQYYPKDKQRHIHITENKLSNCHTYEAPG 1200
Db 1141 lppsniyqqmsttsei y p t n t g p s r s v y s e y q y y p k d k r h i h i t e n k l s n c h t y e a p g 1199
Qy 1201 AKQSSPISSQFASVRRQQLPPNCSIGRESARFKVLNTDQGNQNNLLDLGSSMCYNGLA 1260
Db 1200 akqsspi ssq f a s v r r q l p p n c s i g r e s a r f k v l n t d q g n q n l l d l g s s m c y n g l a 1259
Qy 1261 DSGCGGSPSPMAMLSHEDHALYHTADGLDDMERLYKVDQEPQPPQQQLIPLVPQH 1320
Db 1260 dsgcgsgsp m a m l s h e d h a l y h t a d g l d d m e r l y k v d e q p p q q q l i p l v p q h 1319
Qy 1321 PAEGHLQSWRNQSTRSRKNGQCEIKEPSEL IYAPGVSASERSLLNSGSGTS SQPAGHN 1380
Db 1320 paeghlqswrnqstrsrkngqceikepseliyapgv s a s e r s l l n s g s g t s s q p a g h n 1379
Qy 1381 v 1381
Db 1380 v 1380

RESULT 3

Y13563

ID Y13563 standard; Protein; 1395 AA.

XX

AC Y13563;

XX

DT 30-JUL-1999 (first entry)

XX

DE *Drosophila* Robo 1 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.

XX

OS *Drosophila* sp.

XX

PN W09925833-A1.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24327.

XX

PR 14-NOV-1997; 97US-0065543.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55767.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX
PS Disclosure; Page 30-33; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commisuresless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1395 AA;

Query Match 24.6%; Score 1786.5; DB 20; Length 1395;
Best Local Similarity 30.1%; Pred. No. 1.1e-101;
Matches 441; Conservative 242; Mismatches 507; Indels 275; Gaps 39;

2 ENPRIIEHPMDTTPKNDPFTFNCQAEQNPPTTIQWFKDGRLEKT-DTGSHRIMLPAGGL 60
Db 54 qspriehptdlvkknepatlncckveqkpeptlewfkdgepvstnekkshrvqfkdgal 113
Qy 61 FFLKVIHSRRESADAGTYWCEAKNEFGVARSRNATLQAVLRDEFRLPANTRVAQGEVAL 120
Db 114 ffyrmtgkqgdggyewcvaknrvgqavsrhaslqiaivrddfrvepdkdrvakgetal 173
Qy 121 MECGAPRGSPEPQISWRKNG-----QTLNLVGNKRIRIVDGGNLAIQEARQSDGGRYQC 174
Db 174 lecgpkgipeptliwkdgvplddlkamsfgassrvivdggngllisvepidegnykc 233
Qy 175 VVKNVVGTRESATAFLKVHVRPFLIRGPQNTAVGVSSVVFQCRIGDPLDVLWRRAS 234
Db 234 iaqnlvgtrressyaklivqkvpyfmkpkdqvmlygqtatfhcsvggdppkvkwk-e 291
Qy 235 GGNMPLRKFSLWLSASGRVHLEDRSLKDDTLEDMEYTCADNAVGGITATGILTVH 294
Db 292 egnipvsrarilh-----deksleisnltptdegtyvceahnvqgisaraslivh 342
Qy 295 APPKFVIRPNQNLVEIGDEVLFECQANGHPRPTLYSVVEGSSLLLPYRDRMEVTLTP 354
Db 343 appnftkpsnkkgvngvqvlpcmasgnppsvfwtkegvstlmfnsshgqyva--- 399
Qy 355 EGRSVLSIARFAREDSGVVTCNALNAVGVSSRSTVVSVDQFELPPPIIEQGPVNQTL 414
Db 400 -adgtlqtdvrqdeggyv-csafsvvdsstvrflqvssvderpppiiqigpanqtlp 457
415 VKSIVVLPCTILGTPVPQVSWYLDGIPIDVQEHERRNLSAGALTSIDLQRHEDEGLYTC 474
458 kgsvatlpcratgnpsrikwfhhdgha--vqagnrsiigsslrvidlql-sdsgtytc 514
Qy 475 VASNRNGKSSWGYLRDLTPTNPNIKFFRAPELSTYPGPKQPMVEKGENSVTLWSWRS 534
Db 515 tasgergetswaatltvekggstl--hraadpstypappgtpkvlnvrsstslrwaks 572
Qy 535 NKVGGS--SLVGVIEMFGKNETDGVAVGTQVNTFTQTGLLPVNYFFLIRAENSHG 592
Db 573 qekpgavgiigtyvfyfspdltgtwivaahrvdgtqvtisgltpgtsyvfivraentg 632
Qy 593 LSLPSPMSEPITVGTRYFN--SGLDLSEARSLSGDVVELSNASVVDSTSMKLTQI-- 648
Db 633 isvpsqlsnviktieadfaasandlsaar-tlgtksvelidasainasavrlevmlhv 691
Qy 649 -INGKYVEGFYVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSASASASALISTKPNIA 707
Db 692 sadekyveglrihykdasv----- 711
Qy 708 AAGKRDEGTNQSGGAPTPLNTKYRMLTILNGGASSCTITGLVQYTYLFFIYVFPYKSV 767
Db 712 -----sagyshtvmd-asaesfvvgnlkytkyeffltpffeti 750

Qy 768 EGKPSNSRIARTLEDVPSEAPYGMELLLNSSAVFLKWKAPELKDRHGVLNHYHVIRGI 827
Db 751 eggpsnsktaltiedvpsappdnigimynqtagvwrvtpppsqhngnlygykiev--- 807
Qy 828 DTAHNSRILTNVTIDAASPTLVLANLLEGVMTYVGAAGNAGVGPYCPATLRLDPIT 887
Db 808 -sagntmkvianmtlnatttsvllnnlttgavysvrlnsftkagdgypyskplsfmdp-t 865
Qy 888 KRLDP-----FINQRDH--VNDVLTQPFIIILGAI 916
Db 866 hhvhpprahpsgthdgrhegqdltyhnnnippgdnptthkkttdylsgpwlmlvciv 925
Qy 917 LAVLMLSPG-AMVFVKRKHMMKQ-SALNMRGNHTSDVLKMPSLSARNNGYWLDSSTG 974
Db 926 llvlvisaaismvyfkrkhqmtkelghlsvvsn-----eitalninskeslwidhrg 979
Qy 975 GMYWRPSPGDSLEMQDHIADYAPVCGAPGSPAGGTSSTGGSGAGSGAGSGDDIHGGH 1034
Db 980 ---wr-----tadtdd-----sglsesklshvn 1001
Qy 1035 GSERNQRYVGEYSNIP--TDYAEVSSFGKAPSEYGRHGNASPAYATSSILSPHQOQQ 1092
Db 1002 ssqsn-----yngsdggtadyaevdtrnlrttfncrkspdnptpattmigtsssetc 1054
Qy 1093 QPQRYQRPVPGYGLRPMHPHYQOQQHQOQQAQOQH--QALQHQQLPPSNIYQOM 1150
Db 1055 tkttisadks-gthspysdafagqvpavpvvksnlyqypvepinwseflpp----- 1106
Qy 1151 STTSEIYPTNTGSPRSVYSEQYYPKDKQRHIH-----ITENKLSNCHTYEAPGAKQS 1204
Db 1107 -----ppehppsstygyaqqspessrksksagsgistnqsilnasihsssggfsa 1159
Qy 1205 SPISSQFASVRQQLPP-----NCSIGRESARFKVLNTDQGNQNLDDLGGSSMCY 1256
Db 1160 wgvspqyava---cpenvysnplsavaggtqnryqitptnqhppqlpay----- 1206
Qy 1257 NGLADSGCGSPSPMAMLSHEDEHALYHTA-----DGLDDMERLYVKVDEQPP 1307
Db 1207 --fattgpggavppnhlpfatqrhaaseyqaglnaarcaqsracsdcaldatpmpqpp 1264
Qy 1308 QQQQQLIPLV-----PQHPAEGHLQSWRNQSTRSRKNGQECIKEPSELIYAP 1355
Db 1265 p-----vpvpegwyqpvhpnshpmhptssnhqiy-----qcssecsd--hsr 1304
Qy 1356 GSVASERSLLNSGSGTSSQPAGHN 1380
Db 1305 ssqshkrqlgleehgssakqrgghh 1329

RESULT 4

Y08401

ID Y08401 standard; Protein; 1395 AA.

XX

AC Y08401;

XX

DT 24-JUL-1999 (first entry)

XX

DE Drosophila sp. ROBO1 protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
cell morphology; screening assay.

XX

OS Drosophila sp.

XX

PN WO9920764-A1.

XX

PD 29-APR-1999.

XX

PF 20-OCT-1998; 98WO-US22164.

XX

PR 14-NOV-1997; 97US-0971172.

PR 20-OCT-1997; 97US-0062921.

XX

XX

Comm polypeptide; Robo polypeptide; commissureless; roundabout; modulation; nerve cell function.

PN WO9920764-A1.

```

RESULT 7
Y13565
ID Y13565 standard; Protein; 1297 AA.
XX
AC Y13565;
XX
DT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX

```

```

RESULT      8
Y08403
ID   Y08403 standard; Protein; 1297 AA.
XX
AC   Y08403;
XX
DT   24-JUL-1999 (first entry)
XX

```

```
Qy 439 GIPIDVQEHERRNLSAGALITISDLQRHEDEGLYTCVASNRNKGSSWSGYRLDPTPN 614  
Db 490 glpiditd-srisqhsstgslhiadlkk-pdtgvvyctiaknedgestwsasltvedhts-n 546  
  
Qy 499 IKFFRAPELSTYPGPPGPKQMVEKGENSVTLWTSRNSKNVGSSSLVGVVIEMFGKNETDGW 558  
Db 547 aqfvrmpdpnsfpssptqtplivnvtdevelhwnaptsagagitgyilqyspdlggtw 506  
  
Qy 559 VAVGTRVQNTTFTQTGLLPGVNYFFLIRAENSHGLSLPSPMESEPIVGTTRYFNSGL---- 614  
Db 607 fnipdyvasteyrikglkphsymfviraenekigtpsvssalvttskpaqvalsdkn 666  
  
Qy 615 --DLSEARASLLSGDVVELSNASVVDSTSMKLWQIIN-GKYVEGFVYVARQLPNPIVNN 671  
Db 667 kmdmaiaeakritseqlkleevktinstavrlfwkrkleelidgyyikwr----- 717  
  
Qy 672 PAPVTSTNTPLLGSTSTSASASASASALISTKPNIAAAGRKGDETQSGGGAPTPLNTKY 731  
Db 718 -gpprtndngvvnvntps----- 734  
  
Qy 732 RMLTILNGGGAASSCTITGLVQTYLYEFFIVPFYK---SVEGKPSNSRIARTLEDVPSEAP 788  
Db 735 -----tenyvsnlmpftnyeffviptyghvsihgapnsmdvltaeappslpp 784  
  
Qy 789 YGMEALLNSSAVFLKWKAPELKDRHGVLNLYHVIRGIDTAHNFSRILTNVITIDAASPT 848  
Db 785 edvrirmnlntlrlskwkpkadgingilkqgfivvg--gapnnr---nitteraas 839  
  
Qy 849 LVLANLTGEVMYTIGVAAGNNAAGVGPY--CVPATLRLDLPITKRL-----DPFIN--QR 897  
Db 840 vtlfhltvmtgykirvaarsngggvshgstsevimmqdtlekhlaaqgenesflyglink 899  
  
Qy 898 DHVNVLTPQWFIIILGAILAVLMLSFCAMVFVK-----RKHMMMQSALNTMRGN 948  
Db 900 shvp-----vivivaillfiwfiiaycywrnsrsndgkdrsfikindgsvh-masn 950  
  
Qy 949 HTSDVLKMPS-----LSARNNGYWLDSST-----GGMVWRPSPGGD 985  
Db 951 nlwdvaqpnpqnpmyntagrtmmnrnggalysltpnaqdfinnccdsygmhrpg--- 1006  
  
Qy 986 SLEMQKDHIADYAPVCAGAPSPAGGGTSSGGSGGAGSGSGDDIHGGHGSERNQQRYVC 1045  
Db 1007 ----sehhyhaqltgpgn----- 1022  
  
Qy 1046 EYSNIPTDYAEVSSFPGKAPSEYGRHGNASPAPYATSSILSPHQO-----QQQQQPRYQQ 1099  
Db 1023 -----amstf-----ygnqyhddpspyatttlvlsnqqpawldnkmlrapampt 1066  
  
Qy 1100 RPVPGYGLQRMHP-----HYQQQQHQQQQAQO-----THQQHQALQQHQQLPPSNI 1146  
Db 1067 npvp-----pepparyadhtagrrsrssrasdgrgtlngglihrtsgsqrsdspptdv 1120  
  
Qy 1147 -YQQMSTTSEIYPTINTGPSRSVSEQYYYPKDKQRHITHITENKL-----SN----- 1191  
Db 1121 syvglhssd----gtgsskertgertrpp-----nktlmdfippppsppppg 1164  
  
Qy 1192 CHTYEAPAGA---KQSSPISSQFAS-----VRRQQLPPNCSIGRESARFKVLNTD 1238  
Db 1165 ghvydtatrrqlrgtspredtydsdgafarvdvnarptsrnrlnggrplkgk--rdd 1222  
  
Qy 1239 QGKNQONLLDLGSSSMCYNLGLADSG-----CGGSPSPMAMLMHSHEDEHALYH 1285  
Db 1223 dsqrsslmdddgssseadgenssegdvprggyvrkavprmgiasastla-----hscyg 1274  
  
Qy 1286 t 1286  
Db 1275 t 1275
```

RESULT 9
W83927
ID W83927 standard: Protein: 753 AA.

XX
AC W83927;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human T85 protein.
XX
KW T85; FHMB-6D4; FMHV-SD4; human; neurological disorder; therapy;
KW diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Protein 21..753
FT /label= Mat_protein
FT Region 525..610
FT /note= "has homology to a fibronectin type III domain"
FT Region 638..727
FT /note= "has homology to a fibronectin type III domain"
FT Region 43..101
FT /note= "has homology to a Ig superfamily domain"
FT Region 145..203
FT /note= "has homology to a Ig superfamily domain"
FT Region 237..298
FT /note= "has homology to a Ig superfamily domain"
FT Region 329..394
FT /note= "has homology to a Ig superfamily domain"
FT Region 433..491
FT /note= "has homology to a Ig superfamily domain"
FT Peptide 247..249
FT /note= "RGD motif"
FT Domain 516..600
FT /note= "cytokine receptor homology N-terminal domain"
XX
PN W09848051-A2.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07714.
XX
PR 10-OCT-1997; 97US-0062017.
PR 18-APR-1997; 97US-0044746.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
Holtzman D, McCarthy SA;
DR WPI; 1999-024021/02.
DR N-PSDB; V69278.
XX
PT New isolated human FTHMA-070 and T85 proteins - used to develop
PT products for the diagnosis and therapy of disorders involving
PT cellular processes, e.g. neuronal development.
XX
PS Claim 31; Fig 3; 127pp; English.
XX
CC This is the amino acid sequence of a novel human protein designated
CC T85, and also referred to as FMHB-6D4 and FMHB-SD4. T85 cDNA (see
CC V69278) was identified in a human foetal brain cDNA library using a
CC screen designed to identify genes encoding proteins having a
CC functional signal sequence. T85 nucleic acids and polypeptides of
CC the invention are useful as modulating agents in regulating a
CC variety of cellular processes. They can be used for identifying
CC compounds which bind to or modulate the activity of the polypeptides
CC (claimed). They can also be used in screening assays, detection
CC assays (e.g. chromosomal mapping, tissue typing, forensic biology),
CC predictive medicine (e.g. diagnostic assays, prognostic assays,
CC monitoring clinical trials, and pharmacogenomics), and methods of

CC treatment (e.g. therapeutic and prophylactic) e.g. for neurological
disorders.

XX
SQ Sequence 753 AA;

Query Match 17.3%; Score 1261; DB 20; Length 753;
Best Local Similarity 35.0%; Pred. No. 1e-69;
Matches 285; Conservative 129; Mismatches 284; Indels 116; Gaps 21;

Qy 4 PRIIEHPMDTTPVKNDPFTFNCQABGNPTPTIQWFKDRELKTDG---SHRIMLPAGGL 60
Db 29 privehpsdlivskgepatlnckaeqrptptlewkyggervetdkdprshrmllpsgsl 88
Qy 61 FFLKVIHSRR-ESDAGTYWCEAKNEFGVARSNATLQVAVLRDEFRLEPANTRVAQGEVA 119
Db 89 fflrivhgrksrpddegvyvvarnylgeavshnaslevailrddfrqnpssdmvmvagepa 148
Qy 120 LMECGAPRGSPPEQISWRKNGQTLNLVGNKRIRI-VDGGLAIQEARQSDGRYQCVVK 178
Db 149 vmecqpprghpeptiswkdgspld---dkderitirggklimityrksdagkyvcvtn 205
Qy 179 VVGTRSATAPLKVHVRPFLIRGPQNTAVVGSSVVFQCRIGDPLPDVLRRTASGGNM 238
Db 206 mvgeresevaeltvlerpsfvrkpsnlavtvdssaeafceargdpvptvrwrk---ddgel 263
Qy 239 PLRKFSWLHSASGRVHVLEDRSLKDDVTLEDMEYTCADNAVGGITATGILTVHA--- 295
Db 264 p-----ksryelrddhtlkirkvttagdmsytcvaenmvgaesatltvqvgse 313
Qy 296 PPKFVIRPKNLVEIGDEVLFECQANGHPRTLYWSVEGNSLL---PGYRDMMEVTL 352
Db 314 pphfvvkrpqvvalgrtvtfqceatgnppaifwrrregsnllfsyqppgsrfsvsq 373
Qy 353 TPEGRSVLSIARFAREDSGKVVTCNALNAVGSVSSRTVSV-DTQFELPPPIIEQGPV 411
Db 374 tgd---ltitvgrsvdgyyl-cqtlvagsitkaylevtdviadrpppvirgqpv 428
Qy 412 TLPVKSIVVLPCLRTLGTPVPQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDEGL 471
Db 429 tvavdgtfvlscvatgspvptilwrkdglvstqdsriklen-gvlgir-yaklgdgr 486
Qy 472 YTCVASNRNGKSSWSGYLRD-----TPTNPNIKFRAPELSTYPGPGPKQMEVK 522
Db 487 ytciaastpsgeatwsayievqefgvpqprrptdpnl-----ipsapskpevt 536
Qy 523 GENSVTLSTWTKSNKVGSSLVGYVIMFPGKNETDGVAVGTRVQNTFTTQTLPLGVNYF 582
Db 537 srntvtlsw-qpnlnsgatptsyileafshasgsswtvaenkvtsaiklqpnayl 595
Qy 583 FLIARNSHGLSLSPMSEPI-TVGTRYFNSGLDLSEARASLLSGDVLSNASVVDSTS 641
Db 596 flvraanaygisdpqisdvpktdqlvtsqgvdhkvqre-lgnavlhlhnpvtlssss 654
Qy 642 MKLTWQI-INGKYVEGFYVYARQLPNPIVNNPAPVTSNTNPLLGSTSTASASASALI 700
Db 655 levhwtdvqgsgyiqgykilyr----- 676
Qy 701 STKPNIAAAGKRDGETNQSGGAPTPLNTKYRMLTILNGGASSCTITGLVQYTLYEFFI 760
Db 677 -----psganhgesdlwlvfevrt-----aknsvipdlrkgyvneika 715
Qy 761 VPFFYKSEVGKPSNSRIARTLEDVPSEAPYMEAL 794
Db 716 rpffnefggadseikfaktleegydeafdfhal 749

RESULT 10
R13144
ID R13144 standard; Protein; 1728 AA.
XX
AC R13144;
XX

RESULT 11
R68553
ID R68553 standard; Protein; 1447 AA.
XX
AC R68553;
XX
DT 05-JUL-1995 (first entry)
XX
DE Deleted in colorectal carcinoma (DCC).
XX
KW Tumour suppressor; deleted in colorectal carcinoma; antibody;
KW cancer diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..1063
FT /note= "DCC epitope"
FT Misc-difference 3369..4341
FT /note= "DCC epitope"
FT Misc-difference 26..1126
FT /note= "DCC epitope on extracellular domain"
FT Misc-difference 1123..1447
FT /note= "DCC epitope on intracellular domain"
XX
PN W09428161-A.
XX
PD 08-DEC-1994.
XX
PF 18-MAY-1994; 94WO-US05277.
XX
PR 26-MAY-1993; 93US-0068950.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Bruskin A, Jarosz DE, Johnson K, Kinzler KW, Vogelstein B;
PI Zabrecky JR;
XX
DR WPI; 1995-022830/03.
DR P-PSDB; Q80196.
XX
PT Antibodies specific for tumour suppressor gene product, DCC -
PT useful for detecting expression of DCC gene, for cancer diagnosis.
XX
PS Claim 4; Page 24-28; 39pp; English.
XX
CC The protein represents the DCC tumor suppressor, and epitopes are
CC identified which are used in the generation of polyclonal or
CC or monoclonal antibodies against DCC. The antibodies can detect
CC DCC protein in biological samples (including tumour tissue,
CC peripheral blood mononuclear cells or a tumour biopsy lysate)
CC despite low levels of DCC expression, and are therefore useful in
CC cancer, especially colorectal carcinoma, diagnosis.
XX
SQ Sequence 1447 AA;

Query Match 7.6%; Score 551.5; DB 16; Length 1447;
Best Local Similarity 21.8%; Pred. No. 1.3e-25;
Matches 312; Conservative 177; Mismatches 542; Indels 397; Gaps 64;

Qy 5 RIIHPPMDTTPVKNPFTFNCQAEGR-PTPTIQWFKDGRLEKTDGSHRIMLPAGGLFFL 63
Db 41 rflsepsdavtmrgnvlldcsaegrdgvpikwkkgdghlalmgderkqqlsngslliq 100

Qy 64 KVIHSR-RESAGTYWCEAK-NEFGVARSRNATLQVA-VLRDEFLEPANTRVAQGEVAL 120
Db 101 nilhsrhhkpddeglyqceaslgsdgsiisrtakvavagplr--flsqtesvtafmgdtvl 158

Qy 121 MECGAPRGSPEPQISWRKNGQTLN-LVGNKRIRIVDGGNLAIQEARQSDGRIYQCVKVN 179
Db 121 MECGAPRGSPEPQISWRKNGQTLN-LVGNKRIRIVDGGNLAIQEARQSDGRIYQCVKVN 179

Db 159 lkcev-igepmptihwgknqgdltpipgdsrvvlpssgalqisrlqpgdigyrcsarnp 217
Qy 180 VGTRESATAFLKV-----HVRPFLIRGPQNTAVVGSVVQCRIGGDLPLDVLWRRTA 233
Db 218 assrtgneaevrilsdpqlhrqlyflqrpsnvvaiegkdavleccvsgypppsftwlrge 277
Qy 234 SGGNMLPKRFSWLHSASGRVHVLEDRSLKLDVLTEDMGEYTCADNAVGGITATGILTV 293
Db 278 eviqlrskys-----llgssnllisnvtddsgmytcvvtynkenisasaeltv 327
Qy 294 HAPPKFVIRPKNQLVEIGDEVLFECQANGHPRPTLYWSVEGNSLLLPYRDGRMEVTLT 353
Db 328 lvpwflnhpsnlyayesmdiefectvsgkpvptvnmkngd--vvpv---sdyfqiv-- 380
Qy 354 PEGRSVLSIARFAREDSGKVVTCNALNAVGSVSSRTVVSVDTFELPPPIIEQGPVNQTL 413
Db 381 --ggsnrlilgvksdeg-fygcvaeneagagt-----saqlivpkpaipssslpsa 431
Qy 414 PVKSIVVLPCLTGTTPVQVSWYLDGIPDV-----QEHERR---NLSDAGA 457
Db 432 prdvvpvl---vssrfvrlw---rppaekgnigtftvffsregdnreralnttqpgs 484
Qy 458 --LTISDLQRHEDEGLYT--CVASNRNGKSSWSGYLRDLTPTNPNKFFRAPELSTYGP 513
Db 485 lqltvgnl---kpeamytfrvaynewgpgg-----ssqpkvatqelqvg-pgp 530
Qy 514 PGKPMVEKGENSVTLNRSNKGSSSLVGYVIEFM-----GKNETDGVAVGTRVQN 567
Db 531 venlqavstspsilitweppayang-pvqgy--rlfctevstgkeq-----nievdg 580
Qy 568 TTFQTGTLPGVNYFFLIRAENSHGLSLPSPMSEPITVGTTRYFNSGLDLSEARASLLSGD 627
Db 581 lsykleglkfkfeyslrlaynryg---pgvstdditvt-----lsdypsappgnv 629
Qy 628 VVELSNASVVDSTSMKLTW-----QIINGKYVEGFYV----- 659
Db 630 sle-----vvnrsikvswlpppsgtqng-fitgykirhrktrrgemetlepnnylf 683
Qy 660 -----YARQLPNPIVNNPAP-----VTSNT--NPL---LGSTSTSASASASALIS 701
Db 684 tglekgsqysfqsamtvgntgppsnytaetpendldesqvpdpsslhvrpqtncilm 743
Qy 702 T-----KPNIAAGKRDG-----ET-----NQSG 720
Db 744 swtpplnpivrvyilgygvgspyaetvrvdsqkyryssierlessshyvislkafnng 803
Qy 721 GGAP-----TPL----- 727
Db 804 egvplesatrsitdptdpvdyplldfptsvpdlstplppvgqavalhdavrvs 863
Qy 728 -----NTKYRMLTI--LNGGAS-----SCTITGLVQTYLFFVFPF 763
Db 864 wadnsvpnkqtsevrlytrvrtfsasakyksedttlsytatgkpnmyefsvmvt 923
Qy 764 YKSVGKPSNSRIARTLEDVPSAPYGMALLNNS--AVFLKWKAPELKDRHGVLLNYH 821
Db 924 knrrsstwmtahattyeaapsapkdfvitregkpravivswqpp--leangkitay 981
Qy 822 VIVRGIDTAHNSRILTNV-----TIDAASPTLVANLTGVMYTVGVAAGNAGVG 873
Db 982 l-----fytlknipiddwmetisgdrldthqimldntmyyfrigarngskvg 1031
Qy 874 PYCVP---ATLRLD-----PITKRL-----DPFINQRDHNVDVLT 905
Db 1032 plsdpilfrtlkvehpdkmandqgrhgdgywpvdtalnldrstlneppigqmhphgsvt 1091
Qy 906 QP-----WFIILLGAILAVLMSFGAMVFKRKHHMMQKQALNTRMGNH--TSDVLKMP 957
Db 1092 pqknsnllvliivtvgvitvllvvvavictrrssaqqkrkrathsagkrkysqkdlrpp 1151
Qy 958 SLSARNGNGYWLDSSTGGM--VWRPS---PGGDSLEMQKDHADIAPVCGAGPSAGGGT 1012
Db 1152 dl-----wihheememkniespgtdpagrdsplqs--cgdltpvshsgsetqlgsk 1201

```

Qy 5 RIIEHPMDTTPVKNDPFTFNCQAEGN-PTTIQWFKDGRGLKTDGSHRMLPAGGLFLF 63
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 41 rflsepsdavnrmrgnvlldcsaesdrgvpvikwkkgdihlalmgderkqqlsngslliq 100

Qy 64 KVIHSR-RESDAQTYWCEAK-NEFGVARSRNATLQVA-VLRDEFLEPANTRVAQGEVAL 120
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 101 nilhsrhhkpddeglyqceaslqsgsilstrakvavagplr--flsqtesvtafmgdvtl 158

Qy 121 MECGAPRGSPEPQISWRKNQOTLN-LVGNKRIRIVDGGNLAIQEARQSDDDRQYQCVVKNV 179
   : : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 159 lkcev-igepmptihwknqgdltipgdsrvvvlpsgalqisrlqpgdgigyracsarnp 217

Qy 180 VGTRESATAFLKV-----HVRPFLIRGPQNTAVVGSSVVFQCRIGGDPLDVLWRRTA 233
   : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | :
Db 218 assrtgneaevrilsdpghlrqlyflqrpsnnvaiegkdavleccvsgypssftwlrg 277

Qy 234 SGGNMLRKFSWLHSASGRVHVELEDRSLKDDVTLEDMEYTCADNAVGGITATGILTV 293
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 278 eviglrskkys-----llggsnllisnvtddsgmytcvvtynkenisasaeltv 327

Qy 294 HAPPKFVIRPNQNLVEIGDEVLFECQANGHPRPTLYWSVEGNSLLLPYDRGRMEVTLT 353
   | | : | | | : | | | : | | | : | | | : | : | : | : | : | : |
Db 328 lvppwflnhpsnlyayesmdiefectvsgkpvpvtvnmkngd--vvip--sdyfqiv-- 380

Qy 354 PEGRSVLRIARFAREDSGKVVTNCALNAVGSVSSRTVSVSDTQFELPPPIIEQGVPVQNTL 413
   | | : | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 381 --ggsnlrilgvksdeg-fyqcvaeneagngt-----saqlivpkpaipsssvlp 431

Qy 414 PVKSIVVLPCRTLGTTPVPQVSWYLDGIPDV-----QEHERR-----NLSDAGA 457
   | : | : | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 432 prdvpvl---vssrfvrlsw---rppaekngigtftvffsregdnreralnttqps 484

Qy 458 --LTISDLQRHEDEGLYT--CVASNRNGKSSWSGYLRDLTPTNPNIKFFRAPELSTYPPG 513
   | | : : | : : | | | | : : | | | | | | : : | | | | | |
Db 485 lqltvgnl---kpeamytfrvaynewpgge-----ssqplikvatqelpqv-pgp 530

Qy 514 PGKPQMVKEGNSVTLWSRNSKNVGGSSSLVGYVIEFM-----GKNETDGVWAVGTRVQN 567
   | | | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 531 venlqavstspstilitweppayang-pvqgy--rlfctevstgkeq-----nievdg 580

Qy 568 TTTQTGLLPGVNYFFLIRAENSHGLSLPSPMSEPITVGTFRYFNSGLDLSEARASLLSGD 627
   : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 581 lsykleglkkfteysrlflaynryg---pgvstdditvvt-----lsdvpsappgnv 629

Qy 628 VVELSNASVVDSTSMKLTW-----QIINGKYVEGYFV----- 659
   : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 630 sle-----vvnrsikvslpppsgtqng-fitgykirhrkttrrgemetlepnlnwylf 683

Qy 660 -----YARQLPNPIVNNPAP-----VTSNT--NPL---LGSTSTSASASASALIS 701
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : |
Db 684 tglekgsqysiqvsamtvgtgppsnwytaetpendldesgvdpqpsllhvrpqtnciim 743

Qy 702 T-----KPNIAAGKRDG-----ET-----NMSG 720
   : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 744 swtpplnplnvrgyilgygvgsypaetvrvdskqryysierlessshyvislkafnnaq 803

Qy 721 GGAP-----TPL----- 727
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 804 egvplesatrrsitdptdpvdyppllldfptsvpdlstpmppvgvqavaltdavrsv 863

Qy 728 -----NTKYRLTI--LNGGAS-----SCTITGLVQYTYEFFIVPF 763
   : : | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 864 wadnsvpkngktsevrlytvrrwtsfsasakyksedttlsytatglkptmyefsvmvt 923

Qy 764 YKSVEGKPSNRIARTLEDVPSEAPYCGMEALLNNS--AVFLKWKAPELKDRHGVLNLYH 821
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 924 knrrsstwsmtahattyeaaptsapkdfvtitregkpravivswqp--leangkitayi 981

Qy 822 VIVRGIDTAHNSRILTNV-----TIDAASPTVLANLTEGVMTYVGVAAGNNAWG 873

```

Db 982 1-----fytlidknipiddwimetisgdrllthqimldnldtmyyfrigarnskgyg 1031
Qy 874 PYCVP-----ATLRLD-----PITKRL-----DPFINQRDHVNDVLT 905
Db 1032 plsdpilfrtlkvhpdkmandqgrhgdgywpvdtlnldrstlneppigmqhphgsvt 1091
Qy 906 QP-----WFIILLGAILAVLMLSFQAMVFKRHHMMKQKQALNTMRGNH---TSDVLKMP 957
Db 1092 pqknsnllviivvtvgvitvllvvivavictrrssaqqkrkrathsagkrkgsqkdlrpp 1151
Qy 958 SLSARNNGYWLDSSTGGM--VWRPS---PGGDSLEMQKDHIADYAPVCGAPGSPAGGGT 1012
Db 1152 dl-----wihheememknlekpstgtdpgrdsplgs--cqdltpvshsgetqlgsk 1201
Qy 1013 SSGSGS---GAGSGASGGDDIHGGHSGERNQRY-VGEYSNIPTDYAEVSSFGKAPSEY 1067
Db 1202 stshsgqgteagssmstlerslaarrapraklmipmdagssnp---avvsaipvptles 1258
Qy 1068 GRHGNAPAPYATSSILSPHQQQQQQPRYQORPVP-----GYGLQRPMPHPHYQQQQ 1119
Db 1259 aqypgilpsp---tcgyph-----pqftlrpvpftlsvdrgfgagr----- 1297
Qy 1120 HQQQQAQQTQQHQALQQHQLPPSNIYQQMSTTSEIYPTNTGPSRSV 1167
Db 1298 -----sqsvsegpttqppmlpps---qpehsseapsrtiptacv 1336

RESULT 13

W57900

ID W57900 standard; Protein; 1192 AA.

XX

AC W57900;

XX

DT 25-SEP-1998 (first entry)

XX

DE Protein of clone C0722_1.

XX

KW Human; nutritional supplement; cell proliferation/differentiation;
KW cytokine; immunostimulant; immunosuppressant; haematopoiesis regulator;
KW receptor/ligand activity; cadherin/tumour invasion suppressor;
KW anti-inflammatory; tumour inhibitor; clone C0722_1.

XX

OS Homo sapiens.

XX

PN W09824905.A2.

XX

PD 11-JUN-1998.

XX

PR 05-DEC-1997; 97WO-US22211.

XX

PR 03-DEC-1997; 97US-0984516.

XX

PR 06-DEC-1996; 96US-0762216.

XX

PA (GEMY) GENETICS INST INC.

XX

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX

PI Racie LA, Spaulding V, Treacy M;

XX

DR WPI; 1998-333324/29.

XX

DR N-PSDB; V40887.

XX

PT New isolated polynucleotides encoding secreted polypeptides -
PT isolated from a human foetal kidney cDNA library, a human adult
PT blood cDNA library or a human adult brain cDNA library

XX

PS Claim 36; Page 81-85; 109pp; English.

XX

CC This sequence represents the protein of clone C0722_1, of the
CC invention. DNA encoding this sequence was isolated from a human adult
CC brain cDNA library. The DNAs and proteins can be used as nutritional
CC sources or supplements, or may exhibit cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing

CC activity, haematopoiesis regulating activity, receptor/ligand activity,
CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC tumour inhibition activity or other activities.

XX

SQ Sequence 1192 AA;

Query Match 7.5%; Score 548.5; DB 19; Length 1192;
Best Local Similarity 21.2%; Pred. No. 1.5e-25;
Matches 273; Conservative 157; Mismatches 510; Indels 345; Gaps 48;

Qy 3 NPRIIE-----HPMDTTPKNDPFTFCQAEQNPPTTIQWFKDRELKTDGSH 51
Db 35 dpkilledvqpptitqspkdyidpreniviqceakgpppsfswtrngthfidkdp 94
Qy 52 RIMLPAGGLFFLKI-VHSRRESAGTYWCEAKNEFGVARSNATLQV--AVLRDEFRLP 108
Db 95 vtmkpgtgtliinimsegkaetyegvyqctarnergaavsninnivprsrplwtkeklep 154
Qy 109 ANTRVAQGEVALMECGAPRGSPPEQISWRKNG----- 140
Db 155 it--lsgqslvlpccrppiglpplifwmdnsfqlpqservsglngdyfsnvlpedt 212
Qy 141 -----QTL-----NLVGNKRIR----- 152
Db 213 redyicyarfuhqtqtiqqkpskvvisakssrerptfltpgnasnkkeelrgnvisle 272
Qy 153 -IVDG-----GNLAIQEARQSDDGRYQCVKNVVGTRRES 185
Db 273 claeqlptpiywakedgmlpknrtvyknfektliqihvseadsgnyqciaknalgaihh 332
Qy 186 ATAFKLVHVRPFLIRGPQNTAVGSSVVFQCRIGGDPLPDVLRRTASGGNMLRKFWS 243
Db 333 -tisvrvkaapywitapqnlvlspgedgtlicrangnkp-----risw 375
Qy 246 LHSASGRVHLEDRSLKLDVTL-----EDMGEYTCADNAVGGITATGILTTHAPPK 298
Db 376 ltnvgpieiapddprkldgtliifsnvqerssavvyqcnasneygyllanaafvnlapep 435
Qy 299 FVIRPKNLVEI--GDEVLFECQANGHPRPTLYWSVEGNSLLLPG---YRDGRMEVTL 352
Db 436 riltpantlyqvlarpalldcaffgsplptiewfkagksalhediyvlhengtleipv 495
Qy 353 TPEGRSVLSIARFARESGKVVTCNALNAVGSVSSRTVSV-DTQFELPPP---IEQGP 408
Db 496 -----aqkdstgtytvarnkkgmaknevheikdatwivkqpeyavvrg- 541
Qy 409 VNQTLPVKSIVLPCR-----TLGTPVPQVSWYLDG--IPIDVQEHERRNLSAGALTIS 461
Db 542 -----smvsfeckvkhdtlsl---tvllwkdnrelpsd-----erftvdkdhlvva 585
Qy 462 DLQRHEDEGLITCVASNRNGKSSSGYLRLDTP--NPNIKFFRAPELSTYPPGPKQMV 520
Db 586 dvs-dddsqtytcvanttldsvsasavlsvvaoptp-----ap-vydvnpnpfdlelt 637
Qy 521 EKGENSVTLSTRSNKVGGSSLVGYVIEMFGKNETDGVAVGTRVQNT-TFTQTGLLPV 579
Db 638 dqldksvqlswtpgdd-nnsptkfiieydamhkgplwhqtevsqgttaqlklspyv 696
Qy 580 NYFFLIRABNSHGLSLSPMSEPITVGTIRFNSGLDSEARASLLSGDVVELSNASVDS 639
Db 697 nysfrvmavnsigklspase-----qyltkasepdk-----nptaveg 736
Qy 640 -----TSMKLTWQIINGKYVEG---FYVYARLPNPIVNNPAPVTSNTNPLSGSTSA 690
Db 737 lgsepdnlvitwkpIngfesngpglqykvswrqkdgdewtsvvvanvskiyvsgtptfv 796
Qy 691 SASASASAL--ISTKPNIAAGKRDE-----TNQSGGGAPTPLNT-- 729
Db 797 pylikvqalndmgfapepavvmghsgedlpnvapgnvrnvvnstlaevhwdvpplksir 856
Qy 730 -----KYRMLTILNGGASSCTITGLVQYTYLVEFFIVPFYK 765

```

Qy 4 PRIIEHPMDTIVP---KNDPFTFNQAEAGNPTPTIQWFKDGRLEKTDGTGSHRIMLAGGL 60
   | : | | : | | | | | | | | | | | | | | | | | |
Db 26 piftqephdvifpldlsksevlincaangypsphyrwkqngtdidf-tmsyhyrldggsl 84

Qy 61 FFLKVIHS--RRESDACTYWECAKNEFGVARSRNATLQVAVLREDEFREPANTRVAQGEVA 119
   | : | : | | | | | | | | | | | | | | | | | |
Db 85 ---ainsphtdqdigmyqclatnllgtlslrkaklqfayledfetktrstsvsreggv 140

Qy 120 LMECGAPRGSPEPQISWRKNGQTLNL-VGNKRIRIVDGGNLAIQEARQSDDRYQCVVKN 178
   : : | | : : : | : | : | : | : | : | : | : | : |
Db 141 vllcgppphfgldsyawtfndnplyqvcdnrrfvsqetgnlyiakvepsdvngytcfitn 200

Qy 179 VVGTR-----ESATAFLKVHVRPFLIRGPQNQTAVVGSSVVVQCRIIGDPLDVL 228
   | : | : | : | : | : | : | : | : | : | : | : |
Db 201 keaqrsvsgpptplvqrtgdvmgeyepklevrfpetiqaakdssvklcalfalgnpvdls 260

Qy 229 WRTASGGNMPLRKFSWLHSASGRVHVLEDRS-LKLDVDTLEDMEGYTCEADNAVGGITA 287
   || | : | : | : | : | : | : | : | : | : | : |
Db 261 wrr-ldgsplp-----gkvkyksqailelpnfqgedegfyeciasnlrgnla 308

Qy 288 TGILTVHAPPKFVIRPKNLVEIGDEVLEFCQANGHPRPTLYWSVEG-----N 335
   | : | | : : : : | : | : | : | : | : | : | : |
Db 309 kqqlifyappeweqikqnthlsiydnllweckasgkpnpywtwlkngrlnpeeriqlen 368

Qy 336 SSSL LPG-----YRDGRMEV----- 350
   : | : | : | : | : | : | : | : | : | : | : |
Db 369 gtlitltnlvdsqvyqcaeenkyiiyanaelrvlasapdfskspvkkksfvvggdiv 428

Qy 351 -----TLTPEGR-----SVLSIARFAREDSGVKVTNCNALNAVGS 384
   : : | : | : | : | : | : | : | : | : | : | : |
Db 429 igckpnafpraaiswkrgetlrqskrifledgskliynitrsdags-ytciatnqfgt 487

Qy 385 -----VSSRTVVSVDTFELPPPIIEQGPNVQTLPVKSIVVLPCRTLGTVPVQV--S 434
   | | | : | : | : | : | : | : | : | : | : | : |
Db 488 akntgslivkertvitvp-----pskmdvtvgesivlpcqvhshpsievfvf 534

Qy 435 WYLDGIPID---VQEHERRNLSAGALTISDLQRHEDEGLYTCVASNRNGKSSWSGYLR 490
   | : : | : | : | : | : | : | : | : | : | : | : |
Db 535 wfingdvidlkkgvahferiggesvgdmlmrniqlhh-sgkylctvtq----- 581

Qy 491 LDTPTNPNIKFFRAPELSTYGPPEGKQPMVEKGNSVT---LSWTRSNKVGGSSSLVGVI 547
   : : | : | | | : | : | : | : | : | : | : | : |
Db 582 -----tlelasavadiivrgppgpdpvqvedissttsqlsw-ragpndnspiqifti 633

Qy 548 EMFGKNETD---GWWAVGT-----RVQNTTFTQTGLLPGVNYFFLIRAENSHGLSLPS 597
   : : | : | | | : | : | : | : | : | : | : | : |
Db 634 q-----trtpfsvgwqavatvpellingktyntav-vglspwweyefrvvagnsigigeps 687

Qy 598 PMSE-----PITVGTRYFNSGLDSEARASLLSGDVVELSNASVVDSTSMKLWQII 649
   || : : | : | : | : | : | : | : | : | : | : |
Db 688 epsellrtkasvpvnapvnihggggsrse-----lvitwesi 724

Qy 650 -----NGKIVVEGF-YVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSAS--ASASALIS 701
   | | | | : : | : | : | : | : | : | : | : | : |
Db 725 peelqng--agfgyi-----imfrp-----vgsttwskevsvssvessrfvy 763

Qy 702 TKPNIAAAGKRDEGTNQSGGGAPTPLNTKYRMLTILNGGGAASSCTITGLVQYTLYEFFIV 761
   : | : : | : | : | : | : | : | : | : | : | : |
Db 764 rnesi-----iplspfevkvgvynneg----- 785

Qy 762 PFYKSVEGKPSNSRIARTLEDVPSEAPYGMEEALLNSSAVFLKWKWAPELKDRHGVLNLNH 821
   || | : | : | | | : : | : | : | : | : | : |
Db 786 -----egslstvtivysgedepqlaprgtsglsfsasamevswnaiawnrntgrvlgye 839

Qy 822 VIVRGIDTAHNF--SRILTNTVIDAASPTVLVANLTGVMTYVGVAAGNAGVGPYCVP 878
   | : : | : : | : | | : | : | : | : | : | : |
Db 840 vlywtddskesmigkivrsvngvt-----tknitgkantiyfasvrayntagtgpspp 893

Qy 879 ATLRLDPITKRLDPFPIQRDQHVNDVLTPQ-----WFI-----ILLGAILAVLMLSFGAMVF 929
   : | : | : | : | : | : | : | : | : | : | : :
Db 894 nvv-----ttkkspp-----sqppaniaawkltnsklclnwehvktneneselvlg 937

Qy 930 VYRKHHMMKOSALNTMRGNHSTDVLMKPSLSARNGNGYWLDSSTGGMVWRPSPGGDSLEM 989

```

Db 938 ykilyrqnrqskthiletnntsaellyvpf-----eedyliert-----vsdggdgs 986

Qy 990 QKDHI 994

Db 987 eeiri 991

RESULT 15

W42087

ID W42087 standard; Protein; 1571 AA.

XX

AC W42087;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM2.

XX

KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;

KW signal transduction; trisomy 21; mental retardation;

KW holoprosencephaly; corpus callosum agenesis;

KW schizencephaly; diagnosis; assay; human.

XX

OS Homo sapiens.

XX

PN WO9817795-A1.

XX

PD 30-APR-1998.

XX

PF 23-OCT-1997; 97WO-US19547.

XX

PR 25-OCT-1996; 96US-0029322.

XX

PA (CEDA-) CEDARS SINAI MEDICAL CENT.

XX

PI Korenberg JR;

XX

DR WPI; 1998-271791/24.

XX

DR N-PSDB; V31988.

XX

PT New isolated Down's Syndrome-cell adhesion molecule - used to

PT develop products for detection, diagnosis and therapy of

PT developmental and neurological abnormalities

XX

PS Claim 2; Page 90-95; 109pp; English.

XX

CC This polypeptide comprises Down syndrome-cell adhesion molecule
CC DS-CAM2, an extracellular soluble protein belonging to a novel
CC subclass of the Ig superfamily with highest homology to neural cell
CC adhesion molecules. Its amino acid sequence was deduced from cDNA
CC clones (see V31982) isolated from a trisomy 21 foetal brain library.
CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and
CC lacks the entire transmembrane domain of DS-CAM1. The invention
CC provides human and murine DS-CAM nucleic acid sequences (see also
CC V31981, V31985-87), expression vectors and host cells, transgenic
CC animals, antibodies, antisense oligonucleotides, and primers
CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated
CC with developmental and neurological processes. They can be used in
CC e.g. neural prosthetic devices used in entubulation methods of
CC repairing (regenerating) damaged or severed peripheral nerves, and
CC also in bioassays to identify agonists and antagonists. The products
CC can also be used in detection, diagnosis and therapy of developmental
CC and neurological abnormalities such as Down syndrome, mental
CC retardation, holoprosencephaly, agenesis of the corpus callosum,
CC or schizencephaly.

XX

SQ Sequence 1571 AA;

Query Match 7.4%; Score 538; DB 19; Length 1571;
Best Local Similarity 21.7%; Pred. No. 9.5e-25;
Matches 286; Conservative 166; Mismatches 458; Indels 410; Gaps 55;

Qy 25 CQAEQNPPTPTIQWFKD-----GRELKTDTGSHRIMLPAGGLFLLKVIHSRRES DAGTY 77

Db 246 ckalghpepyrwlkndmplelsgrfqkvtg-----llienlrpsdsgsy 291

Qy 78 WCEAKNEFGVARSRNATLQVAVLRDEFLEPANTRVAQGEVALMECGAPRGSPEPQISWR 137

Db 292 vcevsnrygtakvigrlyvkpk--atisprkvkssvsgsvslscsv-tgtdedqelswy 348

Qy 138 KNGQTLNLVGNKRIRIVDGGNLAIQBARQSDGGRYQCVRK----- 177

Db 349 rngeilnpgknvritginhenlimdhmksdggayqcfvrkdksaqdyvqvvedgtpk 408

Qy 178 -----NVVGT----- 182

Db 409 lisafsekvsvpaepvslmncvkgtpitpitwtldddpilkggshrisgmitsegnvlsy 468

Qy 183 -----RESATAFL---KVHVR-PFLIRGPQNTAVVGSSVVFQCRIG 220

Db 469 lnissqvrddggyrctannasagvlyqarinvgpasirpknitaiagrtdiyhcrvi 528

Qy 221 GDPLPDVLRRTAS----- 234

Db 529 gypysikwyknsnllpfhnrqfafenngtklksdvqkevdegeytcnvlvqplstsq 588

Qy 235 -----GGNMLRKFNSWLHSA---SGRVHLEDR 259

Db 589 vhtvkvppfipgefprfsigqrvfipcvvsgdipi-titwqkdrpipsigvtidn 647

Qy 260 ----SLKLDVTDLEMDGEYTCADNAVGGITATGILTVHAPPKVFIRPKNLVEIGDEV 314

Db 648 idftssrlisnslmhnngnytcianeeaaavehsgqliivrvppkfvvprdqdgdygkav 707

Qy 315 LFECQANGHPRTLYWSVEGNSSL--LLPGYRDGRMEVTLTPEGRSVLSIARFARED SGK 372

Db 708 ilncsaegypptivwfksgagvppqfpialngridiql-----sngsllikhvveedsgy 763

Qy 373 VVTCNALNAVGS-VSSRTVVSVDVQFELPPPIIEQGPVNQTLPVK-SIVVLPCRTLGTPV 430

Db 764 yl-ckvsnvsgadvksmyltvki-----pamitsyp-ntlatqgqgkemsctahgekp 816

Qy 431 PQVSWYLDGIPIDVQEHERR-----NLSAGALTISDLQ----RHEDEGLYTCV 475

Db 817 iivrw-----ekedriinpemarylvstkevgvevistlqilptvredsgffsch 866

Qy 476 ASNRNGKSSWSGLRLDPTNPNIKFRAPELSTYGPQPKQPMVEKG--ENSVTLSWTR 533

Db 867 ainsyged--rgliql-----tvgeppdppeieikdvkartiitrlwtm 907

Qy 534 SNKVGGSSLVGYVIEFPGKNETDGWA-----VGTRVQNTTFTQGLLPVNYFLIRA 587

Db 908 gfd-gnsptgidydec--knksdwsdaqrtdkvsqplnsatiid--ihpsstysirmya 962

Qy 588 ENSHGLSLPSPMSEPITVGTTRYFNSGLDSEARASLLSGDVVLSNASVVDSTSMKLTWQ 647

Db 963 knrigksep---snelti-----tadeaapdgppqev-hlepissqsirvtwk 1006

Qy 648 ----IINGKYVEGFYVYARQLP---NPVNNPAPVTS-----NTNP-----LLGS 685

Db 1007 apkhlqng--lirgyqigreytgggnfqnliisvtdsgdsevytldnlnkftqyglvq 1065

Qy 686 TSTSASASASALISTKPNIAAGKR DGETNQSGGAPTPLNTKYRMLT--ILNG---- 739

Db 1066 acnragtpssqeiitt--tledvpsypennvqaiatspesisiswstlskealngilq 1123

Qy 740 -----GGASCTIT-----GLVQYTLVEFFIVPFYKSVGKPSNSRIAR 778

Db 1124 frviywanlmdgelgeiknittqpsleldgkelytnysiqvlafttragdvrvseqiftr 1183

Qy 779 TLEDVPSAPYGM EALLNSSAVFLKWKAPELKDRHGVLLNHYHVIRGIDTAHNFSRLT 838

Db 1184 tkedvpg--ppagvkaaaasasmfvsw-lpplk--lngiirkytvf-----cshpytpvis 1235

Qy 839 NVTIDAASPTLVLANLTGVMYTVGVAAGNAGVGPYCPVATLRLDPITK---RLDPFIN 895

Db 1236 efeaspdsfsyripnlrnrqysvwwvavtsagrgn--sseilitveplakapariltf-- 1291
Qy 896 QRDHVNDVLTQPWFIILLGAILAVLMLSGAMVFKRKHMMMKQSAINTMRGNHTSDVLK 955
Db 1292 -----sgtvttpw-----mkdivlpckavgdpspavkwkds-----ngtpslvt 1331
Qy 956 MPSSLARNNGYWL-----DSSTGGMVWRPSPGGDS--LEMQKDHADYAPVCGAPGS 1006
Db 1332 idgrrsifngsfliirtvkaedsgyysciannnwgseilnlqvgvppd-----qprl 1385
Qy 1007 PAGGGTSSGSGGAGSGAGGDDIHGGHGSERNQORYVGEYSNIPTDYAEVSSFGKAPSE 1066
Db 1386 tvskttsssitlswlpgdnggssirg-----yilqysednse--qwgspfpispse 1433

Search completed: January 22, 2001, 12:17:58
Job time: 1635 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:24:13 ; Search time 325.28 Seconds
(without alignments)
288.277 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTVPKNDP.....RSLLSNSGSGTSSQAGHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1489.5	20.5	1612	2	T30805	duttl protein - mo
2	1455.5	20.0	1651	2	T14160	transmembrane rece
3	1361.5	18.7	1273	2	T42405	sax-3 protein - Ca
4	1256	17.3	1344	2	T14316	rig-1 protein - mo
5	717	9.9	423	2	T29549	hypothetical prote
6	658	9.0	874	2	T29548	hypothetical prote
7	602	8.3	1028	2	I58164	BIG-1 protein - ra
8	595.5	8.2	1427	2	I51669	tumor suppressor -
9	593	8.2	1232	2	T43027	neural cell adhesi
10	581	8.0	1028	2	A53449	plasmacytoma-assoc
11	579.5	8.0	1277	2	T30532	neural cell adhesi
12	568	7.8	1036	2	S22383	axonin 1 precursor
13	566	7.8	1259	2	A43425	Bravo/Nr-CAM cell
14	565.5	7.8	1272	2	S26180	neurofascin - chic
15	560	7.7	1443	2	I50600	neogenin - chicken
16	558.5	7.7	1268	1	A39640	neural cell adhesi
17	551.5	7.6	1447	2	A54100	tumor suppressor p
18	544.5	7.5	1040	2	A49356	transient axonal g
19	543.5	7.5	1018	2	JC4211	neural adhesion pr
20	543	7.5	1197	2	T30581	neural cell adhesi
21	540.5	7.4	1040	2	A34695	axonal glycoprotei
22	538	7.4	1896	2	T08851	Down syndrome cell
23	536.5	7.4	1018	2	A54744	contactin 1 precu
24	531.5	7.3	1020	2	S05944	neuronal cell surf
25	529	7.3	1239	1	A32579	neuroglian - fruit
26	525.5	7.2	6642	2	T29757	protein UNC-89 - C
27	524.5	7.2	1021	2	A57112	contactin precursor
28	521	7.2	2029	1	TDFFLK	protein-tyrosine-p
29	517.5	7.1	1010	2	JU0094	F11 protein precu

30	517	7.1	1907	2	S50893	protein-tyrosine-p
31	516.5	7.1	1091	2	S01998	contactin precu
32	511.5	7.0	1260	1	S05479	neural cell adhesi
33	509	7.0	1257	1	A41060	neural cell adhesi
34	497.5	6.8	1259	2	S36126	neural cell adhesi
35	495.5	6.8	1897	1	TDHULK	leukocyte antigen-
36	487	6.7	1894	2	C54689	protein-tyrosine-p
37	486	6.7	1898	2	S46216	leukocyte antigen-
38	486	6.7	5175	2	T20992	hypothetical prote
39	485.5	6.7	1912	2	A56178	protein-tyrosine-p
40	484	6.7	5198	2	T43290	hemocentin precu
41	478	6.6	2783	2	T34416	hypothetical prote
42	476.5	6.6	1265	1	A37967	neural cell adhesi
43	475.5	6.5	2222	2	T13924	sdh protein - fruit
44	475	6.5	1256	2	T03096	CD0 protein - rat
45	461.5	6.3	1863	2	S46217	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T30805
duttl protein - mouse
N;Alternate names: transmembrane receptor protein Robol homolog
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30805
R;Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.
submitted to the EMBL Data Library, July 1998
A;Description: The mouse homologue of human DUTTL/H-robol gene: protein sequence
A;Reference number: 220879
A;Accession: T30805
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1612 <WUM>
A;Cross-references: EMBL:Y17793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1
A;Experimental source: brain
C;Genetics:
A;Gene: duttl
A;Map position: 16

Query Match 20.5%; Score 1489.5; DB 2; Length 1612;
Best Local Similarity 28.9%; Pred. No. 2.8e-72;
Matches 402; Conservative 231; Mismatches 516; Indels 243; Gaps 45;

Qy	4	PRIIEHPMDTTVPKNDPFTFNCQAEGNPTPTIQWFKDRELKTDG---SHRIMLPAGGL 60
Db	29	PRIVEHPDLIVSKGEPATLNCKAEGRPPTIEWYKGGERVETDKDDPSRHRMLLPAGSL 88
Qy	61	FFLKVIHSRR-ESDAGTYWCEAKNEFGVARSNATLQAVLRDEFRLPANTRVAQGEVA 119
Db	89	FFLRIVHGRKSRPDEGVYICVARNYLGEAVSHNASLEVAILLRDDFRQNPDMVMVAVGEPA 148
Qy	120	LMECGAPRGSPEPQISWRKNGQTLNLVGNKRIRI-VDGGLAIQEARQSDGRCVQVKN 178
Db	149	VMECQPPRGHPPTISWKDGSPLD---DKDERITIRGKLMITYTRKSDAGRYVCVGTN 205
Qy	179	VVGTRSATFLKVHVRPFLIRGPQNTAVVGSVVQCRIGGDLPLDVLWRRTASGGNM 238
Db	206	MVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRK--DDGEL 263
Qy	239	PLRKFWSLHSGASGRVHLEDRSLKLDVLEDMGEYTCEDNAVGGITATGILTVAHPK 298
Db	264	P-----KSRYEIRDHTLIRKVTAGDMGYSYTCVAMVMGKAESATITVQEPHP 313
Qy	299	FVIRPKQLVEIGDEVLFECQANGHPRTIYWSVEGNSSLLL---PGYRGRMEVTLTP 355
Db	314	FVVKPRDQVVALGRTVTFQCEATGNQPAIFWRREGSQNLFSYQPPQSSRSVSGTGD 373
Qy	356	GRSVLSIARFAREDGKVVTCNALNAVGSVSSRTVVSV-DTQFELPPPIIEQGPVNTLP 414

RESULT 2
T14160

```

Qy      4 PRIIEHPMDTVPKNDPPTFNCQAEGNPPTIQWFKDGRGLKTDTG---SHRIMLPAGGL 60
      |||::||| : : : |||::|||::|||::||| : :|||::|||::|
Db      68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWYKGGERVETDKDDPRSHRMLPSGSL 127

Qy      61 FFLKVIHSRR-ESDAGTYWEACNEKFGVARSNRATLQVAVLRDEFLEPANTRVAQGEVA 119
      |||::||| : : : |||::|||::|||::||| : :||| : |||
Db      128 FFLRIVHGKRKSPDEGVYICVARNYLGEAVSHNASLEVALRRDDFRQNPSDVMVAVGEPA 187

Qy      120 LMECGAPRGSPEPQISWRKNQGTNLVGNKRIRI-VDGGNLAIEARQSDDGRYQCVCVKN 178
      :||| ||| ||| |||::||| : :||| : ||| : ||| : ||| : |||
Db      188 VMECQPPRGHPPTISWKDGSPLD---DKDERITIRGGKLMITYTRKSDAGKYVCVTN 244

Qy      179 VVGTRESAFAFLKVHVRPFLIRGPQNQTAVVSSVVFQCRIGGDPLPDVLRRTASGGM 238
      :||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      245 MVGERESKVADVTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTFGWRK--DDGEL 302

Qy      239 PLRKFSWLHSAAGRVHVEDRSCLKDDVTLEDMEGYTCEADNAVGGITATGILTTHAPPK 298
      | : : : ||| : ||| ||| ||| : ||| : ||| : |||
Db      303 P-----KSRYEIRDHTLIRKVTAGDMGSYTCVAENVMVGKAEASATLTVQEPHP 352

Qy      299 FVIRPKNLVLEIGDEVLFECQANGHPRPTIYWSVEGNSLLL---PGYRGRMEVTLTPE 355
      |||::|||::||| : |||::|||::||| : |||::||| : |||::||| : |||::|||
Db      353 FVVKPRDQVVALGRVTVFQCEATGNPQPAIFWRREGSQNLLFSYQPPQSSSRFSYSQTGD 412

Qy      356 GRSVLSTARFAREDSGKVVTCTNALNAVSGSSVRTVVSV-DTQFLEPPPIIEQGPVNQTL 414
      ||| : ||| : ||| ||| : ||| : ||| : ||| : ||| : |||
Db      413 ---LTVTNVQRSDVGYYI-CQTNLVAGSIITKAYLEVTDVIADRPVPIROGPNQTV 467

Qy      415 VKSIVLPCRTLTGTPVPQVSWYLDGIPIDVQEHERRNLSDAGALTISDLQRHEDEGLYTC 474
      | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      468 VDGTLTLSCVATGSPVPTILWRKDGVLVSTQDSRIKQL-ESGLQLIR-YAKLGDTRYTC 525

Qy      475 VASNRNGKSSWSGYLRLD-----TPTNPNIKFRAPELSTYPGPPGQPMVEKGEN 525
      || : |||::||| : : |||::||| : |||::||| : |||::||| : |||::|||
Db      526 TASTPSCGEATWSAYIEVQEFQVGPVQPPRPRTDPNL-----IPSAPSKPEVTDVSKN 575

Qy      526 SVTLSWTRSNFVGGSSSLVGYVIEFMFGKNETDGVAVGTRVQNTFTTQGLLPGVNYFFLI 585
      :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      576 TVTLLW-QPNLNGSATPTSYIEAFSHAGSSWQTVAEVNKTETFAIRGLKPNAILFLV 634

Qy      586 RAENSHGLSLPSMSEPI-TVGTRYFNSGLDLSEARASLLSGDVS-ELSNASVVDSTSMK 643
      || : |||::||| : ||| : ||| : ||| : ||| : ||| : |||
Db      635 RANAYGISDPQSIDPVKTDVPTTQGVDRHQVQREL--GNVVLHLHNPITLSSSSSVE 692

Qy      644 LTWQI-INGKIVEGFYVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSSASASASALIST 702
      : : : : |||::||| : |||::||| : |||::||| : |||::||| : |||::|||
Db      693 VHWTVDQSQSIQGYKILYR-----NSVVIDLRKGVNYEIKARP 712

Qy      703 KPNIAAAGRDGETNQSGGGAPPLNTKYRMLTIILNGGGASSCTITGLVQYTLYEFFIVP 762
      :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      713 ----PSGASHGSEWLVFEVTRP--TK-----NSVVIDLRKGVNYEIKARP 753

```

Qy 4 PRIIEHPMDTTVPKNDPFTFNCQAEGNPTPTIQWFKDGREL---KTDTGSHRIMLPAGGL 60
| | | | | : | : | | | | : : | | : | | : : | | | | : | |

```

Db      31  PVIEHPIDVVSRGSPATLNCQAKPS-TAKITWYKDGQPVITNKEQVNSHRIVLDTGSL  89

Qy      61  FFLKVIHSR--RESDACTYWCEAKNEFGVARSRNATLQVAVLRDEFLEPANTRVAQGEV  118
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db      90  FLLKVNSGKNKSDAGAYTCVASNEHEGVKSNEGSLKMLAMREDFRVVRPRTVQALGGEM  149

Qy     119  ALMECGAPRGSPPEQIISWRKNGQTLNLVGNKRIRIVDGGNLAIEARQSDDGRYQCVCVK  178
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     150  AVLECSPPRGFPPEPVSWRKDDKELRIQDMPRYTLHSDGNLIIDPVDRSDSGTYQCVANN  209

Qy     179  VVGTRRESATAFLKVHVRPFLIRGPQNQTAVVGSSSVFQCIRIGGDPLPDVLWRRTASGGNM  238
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     210  MVGERVSNPARLSVFEPKFEQEPKDMTVDVGAALVDCRVGTDPQPQITWKR--KNEPM  267

Qy     239  PLRKFSWLHSASGRVHVED-RSLKDDVTLEDMEYTCADNAVGGITATGILTVHAPP  297
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     268  PVT-----RAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPP  317

Qy     298  KFVIRPKNQLVEIGDEVLFECQANGHPRTLYWSVEGNSLLLPGY--RDGRMEVTLTPE  355
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     318  SFQTKPADQSVPAAGGTATFECTLVGQPSPAYFWSKEGQQLLFPYSVADSAGTRKVSPT--  375

Qy     356  GRSVLSIARFAREDSGKVVTCTNALNAVGSVSSRTVVSVDTQF-----ELPPPIIEQGPV  409
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     376  --GTLTIEEVQRVDEGAYV-CAGMNSAGSSLSKAALKVTTKAVTGNTPAKPPPTIEHGHQ  432

Qy     410  NQTLPVKSIVVLP CRTLGTPVPQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDE  469
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     433  NQTLMVGSSAILPCQASGKPTPGISWLRDGLPIDTID-SRISQHSSTLSHIA DLKK-PDT  490

Qy     470  GLYTCVASNRNGKSSWSGYLRDLTPTNPNIKFRAPELSTYGP PGPKPMQVEKGENSVTL  529
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     491  GYVYTCIAKNEDGESTWSASLTVEDHTS-NAQFVRMPDPSPFPSSPTQPIIYVNVTDTEVL  549

Qy     530  SWTRSNNKVGSSLVGYVIEFMGKNETDGWVAVGTRVQNTTFTQTGLLPVGNVYFFLIRAEN  589
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     550  HWNAPSTSAGAPITGYIIQYSPDLGQTFWNI PDYVASTEYRIKGLKPSHSYMFVIRAEN  609

Qy     590  SHGLSLPSPMSEPITVGTRYFNSGL-----DLSEARASLLSGDVVLSNASVVDSTSMK  643
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     610  EKGIGTPSVSVALVTTSKPAQVALSDKNKMDMAIAEKRLTSEQLIKLEEVTINSTAVR  669

Qy     644  LTWQIIN-GKYVEGFYVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSSASASASALIST  702
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     670  LFWKKRLEELIDGYYIKWR-----GPPTNDNQYVNVTSPS-----  706

Qy     703  KPNIAAAGKRDGETNGSGGAPTPLNTKYRMLTILNGGASSCSTITGLVQTYLYEFFIVP  762
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     707  -----TENYVVSNLMPFTNYEFFIVP  727

Qy     763  FYK---SVEGKPSNSRIARTELDVPSEAPYMEALLNNSAVFLKWKAPELKDRHGVLN  819
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     728  YHSVGHSINGAPSNMVDVLTAEPSPLPEDVIRMLNLTIRISWKAPKADGINGILKG  787

Qy     820  YHVIVRGIDTAHNFSRILTNVTIDAASPTLVLANLTGVMYTVGVAAGNNAAGVGPY--CV  877
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     788  FQIVIVG-QAPNNNR---NITTNERAASVTLFHLVTGMTYKIRVAARNSNGGVGVSHGTS  842

Qy     878  PATLRDLPTIKRL-----DPFIN---QRDHVNDVLTQPFIIILLGAILAVLMSFGAMV  928
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     843  EVIMNQDTLEKHLAAQQENESFLYGLINKSHVP-----VIVIVAILIIFVVI IAYC  894

Qy     929  FVK-----RKHMMMKQSALNTRMGNHTSDVLKMPS-----LSARNGNGYWL  969
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     895  YWRNSRNSDGRSFIKINDGSVH-MASNLDVDAQNPNQPNMNYNTAGRM TMNNRNGQAL  953

Qy     970  DSST-----GGMVWRSPGGDSLEMQXDHADIAPVCGAPGSPAGGTS SGG  1016
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     954  YSLTPNAQDFENNCCDYSGITMHRPG-----SEHHYHYAQLTGPGN-----  994

Qy    1017  SGGAGSGASGGDDIHGGHGSERNQQRVYGEYSNIPTDYAEVSSFGKAPSEYGRHGNASPA  1075
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     995  -----AMSTF-----YGNQYHDDPS  1009

```

```

RESULT      5
T29549
hypothetical protein ZK377.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29549
R;Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid ZK377.
A;Reference number: Z20639
A;Accession: T29549
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-423 <NHA>
A;Cross-references: EMBL:U88183; PIDN:AAB52658.1; GSPDB:GN00028; CESP:ZK377.3
A;Experimental source: strain Bristol N2; clone ZK377
C;Genetics:
A;Gene: CESP:ZK377.3
A;Map position: X

```

```

Qy 400 PPPHIEQGPVNQTLTPVKSSIVLPCRTLGTPVPQVSWYLDGIPIDVQEHERRNLSAGALT 459
    ||| ||| |||| | | :|||: | | | :|| ||:|||: | : | :|
Db 28 PPPTIEHGHQNTLMWGGSSAILPCQASGKPTPGISWLRDLGLPIDITD-SRISQHSQSLH 86

Qy 460 ISDLQRHDEGLYTCVASNRNGKSSWSGYLRDLTPTNPNIKFFRAPELSTYGPQPGKPM 519
    |:||: | | :|||: | :|:| | | : | : | : | : | : | : | : | : |
Db 87 IADLKK-PDTGYVTCIAKNEDGESTWSASLTVEDHTS-NAQFVRMPDPSNFPSSPTQPII 144

Qy 520 VEKGENSEVTLWSWTRSNKVGGSSLVGYVIEMFGKNETDGWVAVGTRVQNTTFTQTLPGV 579
    | : ||| | : | : ||:| | : | : | : | : | : | : ||
Db 145 VNVDITVELHWNAPSTSGAGPITGIIQYSPDLGQTWFPIDPYASTEYRIKGLKPSH 204

```

```
Qy 580 NYFFLIRAENSHGLSLPMSPEITVTRYFNSGL-----DLSEARASLLSGDVVELSN 633
   :| |:|||| | : || | :| | :| :| | :||
Db 205 SYMFVIRAENEKGIGTPSVSSALVTTSKPAQVALSDKNKMMDMAIEAKRLTSEQLIKLEE 264

Qy 634 ASVVDDSTSMKLTIQIIN-GKYVEGFYVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSASA 692
   :||::|| :| : ::||: | | :| | : | ||
Db 265 VKTINSTAVRLFWKRRKLEELIDGIYYIKWR-----GPPRTNDNQYVNVTSPS--- 311

Qy 693 SASASALISTKPNIAAAGKRDEGTNQSGGGAPTPLNTKYRMLTILNGGASSCTITGLVQ 752
   : : : :|
Db 312 -----TENYVVSNLMP 322

Qy 753 YTYLEFFIVFPYK----SVEGKPSNRARIATLEDVPSEAPYGM EALLNSSAVFLKKWKAPE 809
   :| |||::|| :| : ||| | | | | : :|| : : |||:
Db 323 FTNYEFFVIPYHSGVHSIHGAPSNSMDVLTAEPSPLPEDVRIRMLNLTTLRI SWKAPK 382

Qy 810 LKDRHGVLNNYHVIRVGIDTAHNFSRLTNVTIDAASPTLVLANLTGEVMTYVGVAAGNN 869
   :|| : : :| :| :| :| : :|| :| :| :| :| :|
Db 383 ADGINGLKGFQIVIVG--QAPNNNR---NITTNERAASVTLFHLVTGMTYKIRVAARSN 437

Qy 870 AGVGPY--CVPATLRDLPITKRL-----DPFIN---QRDHVNDVLTPQPWFILLGAILA 918
   ||| : : :| :| :| :| :| :| :| :| :| :|
Db 438 GGVGVSHGTSVEIMNQDTLEKHLAAQQENESFLYLINKSHVP-----VIVIVAILI 489

Qy 919 VLMLSFGAMVPVK-----RKHHMMKQSALNTMRGNHTSDVLK MPS-----L 959
   : : | :| :| :| :| :| :| :| :| :| :| :|
Db 490 IPFVIIAIYCYNRSRNSDGDKRSFIKINDGSVH-MASNLDWAQNPNQPNMYNTAGRM 548

Qy 960 SARNGNCYWLDSST-----GGMVWRPSPGGSLEMOKDHIADYAPVCGAPGS 1006
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 549 TMNNRNQGALYSUTPNAQDFFNCCDYSGTMHRPG-----SEHHYTHAQLTG GP GN 599

Qy 1007 PAGGGTSSGGSGGAGSGASGGDDIHGHGGERNQRYVGEYSNIPTDYAEVSSF GKAPSE 1066
   : :| :| :| :| :| :| :| :| :| :| :| :|
Db 600 -----AMSTF----- 604

Qy 1067 YGRHGNASPAYATSSILSPHQO-----QOOQQPRYQRPVPYGGLQRP MHP-----HY 1113
   || :| :| |||::|| :| :| :| :| :| :| :| :| :| :|
Db 605 YGNQYHDDPSPYATTTLVLSNQQPAWLNDKMLRAPAMPTNPVP-----PEPPARYADHT 658

Qy 1116 QOOQHQQOOQAQQ-----THQHQALQQHQQLPPSNI-YQQMSTTSEIYP TNGPSRS 1160
   : : :| :| :| :| :| :| :| :| :| :| :| :|
Db 659 AGRRSSRSRASDGRGTLNGLLHHRITSGSQSRDSPPHTDVSYVOLHSSD-----GTGSSKE 713

Qy 1167 VI SEQYYYPKDKQRHIHITENK-----SN----CHTYEAPGA---QOSSPISS 1209
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 714 RTGERRTP-----NKTLMDFIPPPSPNPPPGGHYDTATRQLNRGSTPRED 762

Qy 1210 QFAS-----VRRQQLPPNCISIGRESARFKVLNTDQGNQQNLLDLLDGSSMCYNGL 1259
   :| :| :| :| :| :| :| :| :| :| :| :|
Db 763 TYDSVSDGAFARVDVNARPTSRNRNLGGRPLKGK--RDDDSQRSLLMDDDDGGSSEADGE 820

Qy 1260 ADSG-----CGGPSPMAMLMSHEDEHALYHT 1286
   | | | :| :| :| :| :| :| :| :| :| :|
Db 821 NSEG DVPRGGYRKAVPRMGISASTLA-----HSCYGT 852
```

```

RESULT 7
I58164
BIG-1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C:Accession: I58164
R:Yoshihara, Y.; Kawasaki, M.; Tani, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori
Neuron 13, 415-426, 1994
A>Title: BIG-1: a new TAG-1/F3-related member of the immunoglobulin superfamily with
A:Reference number: I58164; MUID:94338697
A:Accession: I58164
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1028 <RES>
A:Cross-references: EMBL:U11031; NID:q563132; PIDN:AAA63607.1; PID:q563133

```


Db 479 TSQPISLQITVGNLTPEETYN-FRVVAYNEWGPGE-----SSQEVKVVTPQLQV 527
Qy 510 YPGPPGKPMVEKGENSEVTLSTWTRSNKVGSSLVGYVI---EMFGKNETDGVAVGTRVQ 566
Db 528 -PGPEVENLQVVSTAPTSVLISWDPPAYANG-PVQGYRLPCAETFSGREQN-----IEVD 579
Qy 567 NTTFTQTGLLPGVNYFFLIRAENSHGLSLPSPMSEPITVGRYFNGLDSEARASLLSG 626
Db 580 GIVIRLEGLRKFTEYSIRVLAYNRYG---PGVSSEETVVT-----LSDVPSAMPQN 628
Qy 627 DVVELSNASVVDSTSMKLTW-----QIINGKYVEGFV----- 659
Db 629 VSLEVAN-----SRSIKVSWLPPPGTQNG-FITGYKIRHRKTRRGELETLEPNLWYL 682
Qy 660 -----YARQLPNPIVNNPAP---VTSNT--NPLLGSTSTASASASALIST-- 702
Db 683 FTGLEKGSQYSFQAAMTVNGTGPSSDNYTAETPENDLDESQVDPQSSLHVRPLTTSII 742
Qy 703 -----KPNIAAGKRDG-----ET-----NQS 719
Db 743 MSWTPPLNPNIVVGYIIGYGVGSPTAETVRVDSKQRYYSIENLEPSSHYVISLAFNNA 802
Qy 720 GGGAP-----TPL-----NTKYRM 733
Db 803 GEGVPLYESATTRSQTPMDSTMPLPPVGVQAVALTHDAVRVSWADNSVLKNQKTEVRF 862
Qy 734 LTI--LNGGGAS-----SCTITGLVQYLTIEFFIVPFYKSVGKPSN---SRI 776
Db 863 YTIWRWTSYASSKYKADTSLSHVITGLKPNMTYEFVSM---VTGKRSSWSTMAH 918
Qy 777 ARTLEDVPSAPYGMALLNSS--AVFLKWKAPKLDHRGVLLNYHVIRGIDTAHNS 834
Db 919 ATTYETAPTAPKDLTVITRERKPRAVIWSQPP--IEANGKIIDF-ILFYTLKDLQLD 975
Qy 835 R-ILTNVTIDAASETLVLANLTGVMYTVGVAAGNNGVGPVCPVTLRLDPITKRLDPF 893
Db 976 DWIMVTITGDRLTIELDLNL--DTAYYFRIQARNAGKGLPSEPIFR-TPKVEHPDKM 1032
Qy 894 INQRDHVD-----VLTPQ-----WFIILLGAIL 917
Db 1033 ANDQGRHGDGGYNSVDNLIDRSSLENEPIGQMHPHGSVTPQKNSNLLVIVTVGAI- 1091
Qy 918 AVLMLSFGANVFKRKHMMKQKQALNTRMGNH--TSDVLKMPSLSARNGNGYWLDSSTGG 975
Db 1092 TILVVIVIVICTRSSAQQRKRRATHSAGKRGSKQDLRPPDL-----WIHHEEME 1143
Qy 976 M--VWRSPGGDS--LEMOKDHIADYAPVCGAPGSPAGGCTSSGGSGGASGSGGDDIH 1031
Db 1144 MKNIEKPS-GSDTQGRDSPRQSCDITPVSHSQSESQSGKSTSHSG-----PDADVEG 1196
Qy 1032 GGHGS-ERN--QRTYGEYSNIPTDYAEVSSFGKAPSEYGRHGASAPATYSS---IL 1084
Db 1197 SNISTLERTLAARRATRAKLMIPMD-----SQPSSNPVVSAPVPTLESAPYGPIL 1248
Qy 1085 S-----PHQQQQQQQPRYQORVPV-----GYGLQRPMPHYQQQQHQQQQAQTHQ 1130
Db 1249 PSPTCCGYPH-----PQFTLRPVFPPLTVDRGFGTSRVTEVPASQSSSVLSHPQPEH- 1300
Qy 1131 QHQAQQHQQQLPPSNIIQQMSTTSEIYPTNTGPSRSV 1167
Db 1301 -----STSEDAPSRTIPTACV 1316

RESULT 9

T43027

neural cell adhesion molecule L1 - goldfish

N:Alternate names: E587 antigen

C:Species: Carassius auratus (goldfish)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000

C:Accession: T43027

R:Giordano, S.; Laessing, U.; Lottspeich, F.; Stuermer, C.A.O.

submitted to the EMBL Data Library, April 1996

A;Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule

A;Reference number: 222294

A;Accession: T43027

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1232 <GIO>

A;Cross-references: EMBL:U55211; NID:g1305526; PID:g1305527; PIDN:AAA99159.1

C;Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homolog

C;Keywords: cell adhesion; membrane protein

Query Match 8.2%; Score 593; DB 2; Length 1232;
Best Local Similarity 23.2%; Pred. No. 3.7e-24;
Matches 296; Conservative 146; Mismatches 495; Indels 340; Gaps 51;

Qy 4 PRIIEHPMDTTPKNDPPTFNCQAEQNPPTIQTQWFKDRELKTDGSHRIMLPAGGLFFL 63
Db 15 PIITVQPVSHTAFLSDVILACEASGDPAPSFWRWKDGKEFK-----RELLSSGTL-- 65
Qy 64 KVIHSRRES--AGTYWCEAKNEFGVARSNRATL---QVAVLRDEFLEPANTR-VAQGE 117
Db 66 -TAEDKEELHPIQGSRYCYVMS-LGTAVSDLAQLITEPIPTLAKEKR---QKTRSFEEGD 12C
Qy 118 VALMECGAPRGSPEPQISWRKNGQTLNLVGNKRIRI-VDG---GNLAIQEARQSDGGRY 172
Db 121 SAVLYCNPPKSSVTPKIHV-MDMHWRHIPLNERVTTSLDGNLYFANLVNDSRED---Y 175
Qy 173 QC---VKNVVGTRRESATFLKVHVRP---FLIRGPQNG-----TAVVGSSVVF 215
Db 176 TCNAHIINASVILPKER---ISISVTPSNSVLKNRRPQLQKPAAGSHSSYLVRGQTLTL 231
Qy 216 QCRIGGDPLPDVLRRTASGGMPLRKFSWLHSASGRVHVLE-DRSLKLDVLEDMGEY 274
Db 232 ECIPEGLPTPEVQWERMDS---PL-----SPARVRLKYRNLQIESVSEADDEY 279
Qy 275 TCEADNAVGGITATGILTVHAPPKVFIRPKNLQVEIGDEVLFECQANGHPRTLYSVEG 334
Db 280 TCTAQNSQSGVKHHYAVTVEAAPYWRTPENHLYAPGETVRLDQCAEGIPTNITWSM-- 337
Qy 335 NSSLLPGYRGRMEVTLTPEGRSVLSIARFAREDGKVVTCNALNAVGSVSSRTVVSVD 394
Db 338 NGAPIAGTDPPRRHVS---SGTLILTVDVQIS---DTAVYHVEATNKHGNILINHVHV- 390
Qy 395 TQFELPPPIEQGPVNQTLPVKSIVVLPCTRITLTPVPQVSWYLDGIPIDVQEHERRNLS 454
Db 391 --VELPQILTEDDLKYEATEQVTLVLCRTFGSPQPKVDWQITNSGALANAKMSQTS 448
Qy 455 AGALTISDLQRHEDEGLYTCVASNRNGKSS----- 484
Db 449 -GNLQISDVS-EEDSSMYTCSYSTSNMSIAELVVLNRTKIVDPQDLRLVRGDDAVLQC 506
Qy 485 -----WSGYRLRDLTPNPNKIFRAP-----ELST-- 509
Db 507 RYTVDHMLKQPTIQWKKDKHKITSSANDKYTESPDGSLKITDVQMEDSGIYSCEISTKL 566
Qy 510 -----YPGPPGKPMVEKGENSEVTLSTWTRSNKVGSSLVGYVIEMFGK-NET 555
Db 567 DVSATGSIVVLDPKGPSHSELSEKERSVTLWMPGAE--NNSPISEYVIERKEKNPG 625
Qy 556 DG-WVAVGTRVQNTFTTQTGLLPGVNYFFLIRAENSHGLSLPSPMSEPITVGRYFNGL 614
Db 626 KGHWEERYRVPQDITHLEIHLQYISTYHFRVRGVNGIMSESPSPSEYST----- 676
Qy 615 DLSEARASLLSGDVVELSNASVVDSTSMKLTWQIINGKYVEG---FYVYARQ----- 663
Db 677 --PAKPDMPNENVTSVS-----TDSNLVITWQELQRFNGPGFKYKIYVRQEGDSHW 730
Qy 664 ---LPNP--IVNNPAP-----VTSNTNPLLGSTSTASAS 693
Db 731 ESSASNPPIVEGPGTFIPQIKVQAVNELGAGPEPDAEIGSGEDLPLEAPSSVAVSEL 790
Qy 694 ASASALISTKP--NIAAG-----KRDGETNQSGGAPPLNTKYRMLTILNGGA 742

```

RESULT 11
T30532
neural cell adhesion molecule L1 homolog - Fugu rubripes
C;Species: Fugu rubripes
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T30532 ;
R;Riboldi Tunniciiffe, G.R.; Platzer, M.; Nyakatura, G.; Elgar, G.S.; Brenner, S.; Ro
submitted to the EMBL Data Library, September 1997
A;Description: Analysis of the genomic loci of Fugu rubripes homologs of the human di
A;Reference number: Z20848
A;Accession: T30532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1277 <RIB>
A;Cross-references: EMBL:AF026198; NID:q3098263; PID:q3098264; PIDN:AAC15580.1

```

C:Genetics:

A;Introns: 42/1; 47/1; 81/2; 149/1; 190/1; 247/1; 285/2; 347/1; 391/1; 440/1; 477/2; 531/2

A;Note: L1-CAM

Query Match 8.0%; Score 579.5; DB 2; Length 1277;
Best Local Similarity 23.3%; Pred. No. 2.1e-23;
Matches 240; Conservative 139; Mismatches 429; Indels 221; Gaps 35;

Qy 4 PRIIEHPMDTTVPKNDPPTFNCQAEAGNPTPTIQWFKDGRKELKDTGSH-RIMLPAGGLEFF 62
| | | | : | | | | | | | : : | | |
Db 51 PAITTPQESPTVFSVEDLVMRCEASGNPSPTFHMTKDGEEFDPSDDPEMKVTEAGSSVF 110

Qy 63 LKVIHS--RRESAGTYWCEAKNEFGVARSRNATLQVAVLRDEFRLPANTRVAQGEVAL 120
: : : : | | | | | | | : : : :
Db 111 YTLSTMTDLKQYQKICYASNELGTAVSNAVLMDAPPVQKEKKVTECAEAGHSIA 170

Qy 121 MECGAPRGSPEPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDG-GRYQCVVKNV 179
: | : | : | : : : : : | : : : :
Db 171 LSCNPPQSSMQPIIHWMDN-RLRHIRLSDRVVGVKGNLYFANLLTEDSRNDYTCNIQYL 229

Qy 180 VGTRESATAFLKVHVRPFLI----RGQP-----NQTAVVGSVVFCRIGGDPLPD 226
| : | : | : | | : | | : | : : : | : | :
Db 230 ATRITIAKEPITLTVNPSNLVPRNRPPQMRRTGSHSTHALRGQTELELCIVGLPTPK 289

Qy 227 VLWRRITASGGNMPLRKFSWLHSASGRVHVLEDRSLKDDVTLEDMEYTCADNAVGGIT 286
| | | : | : | : | : : | : : : : | : | :
Db 290 VSWLR--KDGEMSESIS-----KDMFDRRLQFTNISESDGGEYQCTAENVQGRTF 338

Qy 287 ATGILTVHAPPKFVIRPKNLQVEIGDEVLFECQANGHPRTLYWSVEG---NSSLLPGY 343
: | : | : | : | : | : | : | : | : | : | :
Db 339 HTYTVTVEASPYWTNAPVSQLYAPGETVKLDCQADGIPSPITITVWNGVPLSALSLEP-- 396

Qy 344 RDGRMEVTLTPEGRSVLSIARFAREDSKVVTCNALNAVGSVSSRTVVSVDTFELPPPI 403
: | : | : | : | : | : | : | : | : | : | :
Db 397 -----RRSLTESGSLILKDVIFG---DTAIYQCAQSNKKGITLANTVYV---IELPPQI 445

Qy 404 IEQGPVNTLPVKISIVLPCRTLGTPVPQVSWYLDGIPDVQHEHRRNLSDAGALTISDL 463
: : : : | : | : | : | : | : | : | : | :
Db 446 LTENGNTYTFVEGQKALLECECTFGSPKPKVTWESSISILLAD-PRVNLNTNGGLEIANV 504

Qy 464 QRHEDEGLYTCVASNRN-----GKSSWSGYRLDTP--T 495
: | : | : | : | : | : | : | : | : | : | :
Db 505 S-HDDEGIYTCLVQGSNISVNAEVEVLNRTVILSPQALRLQPGKTAIFTCLYTDPKLS 563

Qy 496 NPNIKF-----FRAPEL-----STY- 510
: | : : : | : | : | : | : | : | : | :
Db 564 SPLQWRKNDQKIFESHSDKKYTFDQPLIISNVEPDEGVYTCQIITKLDMEASSTLT 623

Qy 511 ----PGPPGKPMVEKGENSVLTSWTRSNKVGSSLVGYIEMFGKN-ETDGVWVAGTRV 565
| | | : | : | : | : | : | : | : | :
Db 624 LCDRPDPVHLQVTNAKHVVTLNWTGDD-NNSPILEYVVEFEDQDMKENGWELKRV 682

Qy 566 QNTFTQTGLLPVNYFFLIRAENSHGLSLPSPMSEPITVGTIRYFNSGLDSEARASLLS 625
: | : | : | : | : | : | : | : | : | :
Db 683 ADKKHVNPLWPMSYRFRVIAINDQKSDPSKLS-----DLYKTPADAPD 728

Qy 626 GDVVELSNASVVDSTSMKLTWQII----NG---KY----- 653
: : : | : : : | : : | : | :
Db 729 SNPEDVRSES-TDPTLVITWEEMDKRNFNGPDFKLYVMWRRVVGSGPDWHEEYTIAPPF 787

Qy 654 ----VEGFYVYARQL-----PNPIV-----NNPAPVTSNTNPLLGSTST----- 688
: | : : : | : | : | : | : | : | : | :
Db 788 IVDVQNFSAFEIKVQAVNKKGLGPEPDPPIIGYSGEDVLEAPLNLGVLEENSTIRVTW 847

Qy 689 -SASASASALISTKPNIAAGKRDEGTQSGGGAPTPLNTKYRMLTILNGGASSCTI 747
: : : : | : : : | : | : : : : | :
Db 848 SAVDKETVRGHLGKYIYLTWGHNRNSR-----AQEPEN---IVMVQTGANEEKSI 896

Qy 748 TGLVQYTYLFEFFIYFFYKSVGKPSNSRIARTLEDVPSAPYGMALLNSSAVFLKWK 807
| | | : | : | : | : | : | : | : | :
Db 897 TNLRPYCHYDLAISAFNSKGEGPLSEKTSFMTPEGVPG-PPMSMQMTSPSEETILHTWP 955

Qy 808 PELKDRHGVLLNYHVIVRGIDTAHNFSLTNVTIDAASPT---LVLANLTEGVMTYGV 864
| : | : | : | : : : | : | : | : | :
Db 956 P--SKPNGILLGVSLYQKMQSDDNPLQV-----VDIASPEITHLTLKGLDRHSYQFL 1008

Qy 865 AAGNNAGVG 873
| | |
Db 1009 MARTAAGKG 1017

RESULT 12
S22383
axonin 1 precursor - chicken
N;Alternate names: neural cell adhesion molecule AxCAM
C;Species: Gallus gallus (chicken)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jan-2000
C;Accession: S22383; S34107; S69332; S22128
R;Zuellig, R.A.; Rader, C.; Schroeder, A.; Kalousek, M.B.; von Bohlen und Halbach, F.
Eur. J. Biochem. 204; 453-463, 1992
A;Title: The axonally secreted cell adhesion molecule, axonin-1. Primary structure, i
A;Reference number: S22383; MUID:92174898
A;Accession: S22383
A;Molecule type: mRNA
A;Residues: 1-1036 <ZU1>
A;Cross-references: EMBL:X63101; NID:g62852; PIDN:CAA44815.1; PID:g62853
A;Accession: S34107
A;Molecule type: protein
A;Residues: 29-49;51-80;84-95;100-117;120-128;130-141;143-176;243-254;256-296;303-336
R;Giger, R.J.; Vogt, L.; Zuellig, R.A.; Rader, C.; Henahan-Beatty, A.; Wolfer, D.P.;
Eur. J. Biochem. 227; 617-628, 1995
A;Title: The gene of chicken axonin-1. Complete structure and analysis of the promote
A;Reference number: S69332; MUID:95172044
A;Accession: S69332
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1002-1036 <GIG>
A;Cross-references: EMBL:X79607
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
C;Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology
C;Keywords: cell adhesion
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-1036/Product: axonin 1 #status predicted <MAT>
F;336-392/Domain: immunoglobulin homology <IMM>

Query Match 7.8%; Score 568; DB 2; Length 1036;
Best Local Similarity 23.1%; Pred. No. 6.4e-23;
Matches 250; Conservative 116; Mismatches 352; Indels 364; Gaps 40;

Qy 4 PRIIEHPMDTTVPK---NDPPTFNCQAEAGNPTPTIQWFKDGRKELKDTGSHRIMLPAGGL 60
| | | | : | : | : | : | : | : | : | :
Db 32 PVFEEQPAHTLFPEGSAAEKVLTCLRARANPPATYRKWMNGTELKMGPDSS-RYRLVAGDL 90

Qy 61 FFLKVIHSRRRESAGTYWCEAKNEFGVARSRNATLQVAVLRDEFRLP----- 107
: : : | : | : | : | : | : | : | :
Db 91 VISNPVKAK---DAGSYQCVATNARGTVVSREASLRFGFLQ-EPSAEERDPVKITEGWGV 146

Qy 108 ----- 107
Db 147 MFTCSPPPHYPALSYRWLLNEFPNFIADGRRFVSQTGNLYIAKTEASDLGNYSFATS 206

Qy 108 -----PANTRVAQGEVALMECCGAPRGSPEPQ 133
| : | : : | : | : | : | : | :
Db 207 HIDFITKSVFSKFSQSLAAEDARQYAPSIKAKFPADTYALTGMVTLCEFA-FGNPVPQ 265

Qy 134 ISWRK---NQQLNLVGNKRIRIVDGGNLAIQEARQSDGGRYQCVVKNVGTRESATAFLK 191
| | | : : : : | : | : | : | : | : :
Db 266 IKWRKLDGQSQTSKWLSSEPL-----LHIQNVDFEDEGTYECAENIKG-RDTYQGRIT 317

Qy 192 VHVRFPLIRGPNQNTAVVGSVVFCRIGGDPLPDVLRRTASGGNMPLRKFSWLHSASG 251
: | : : : | : | : | : | : | : | : :
Db 318 IHAQPDWLDVITDTEADIGSLRNSCVASGKPRAPVWRLR----DGQPL-----ASQN 366


```

RESULT 13
A43425
Bravo/Nr-CAM cell adhesion molecule L1 homolog - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A43425
R;Kayyem, J.F.; Roman, J.M.; de la Rosa, E.J.; Schwarz, U.; Dreyer, W.J.
J. Cell Biol. 118, 1259-1270, 1992
A;Title: Bravo/Nr-CAM is closely related to the cell adhesion molecules L1 and Ng-CAM and
A;Reference number: A43425; MUID:92381110
A;Accession: A43425
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid; protein
A;Residues: 1-1259 <KAY>
A;Experimental source: cerebellum
A;Note: sequence extracted from NCBI backbone (NCBIP:112026)
C;Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; I
F;237-294/Domain: immunoglobulin homology <IMM>

```

```

Qy      8 EHPMDTTVPFANDPFTFNCAQEGNPTPTIIQWFKDGRLEKTDTSGRHIMLPAGGLFFLVKH 67
      : | | | : : : | : | | | : : | : | | : : :
Db     22 QSPKDYIIVDPRENIVIQCEAKGKPPPSFSWTRNGTHFDIKDKAQVTMKPNSGTLVWNIM 81

Qy     68 S-RRESDAQTYWCEAKNEFGVARSNATL--QVAVLRDEFRLPEANTRVAQGEVALMECG 124
      : | : | | | : | | | | : : : | : | | : | : : |
Db     82 GKVAEAYEGVYQCTARNEFGAAISNNIVIXXSXSPLWTEKKELEPNHVR--EGDSLVLNCR 139

Qy    125 APRGSPEPQISWRKNQOTLNLVGNKRIIRDGNGLAIQEARQSDGGR--YQCVVK--NVV 180
      | | | | : : : | : : : | : | : | : | : : :
Db    140 PPVGLPPPIIFWMDNA-FQRLPQSERVSQGLNGDLYFSNV-QPEDTREDYICYARFNETI 197

Qy    181 GTRESATAFLKV-HVRPFLIRGP-----QNQTAVVGSSVVFQCRIGDPLPDVLWR 230
      : : | : | : | : | : : : | : : : | : | : |
Db    198 QQKQKQPISVKVFSTKPVTERPPVLLTPMGSTSNKVELRGVNLLECIAAGLPTPVIRWI 257

Qy    231 RTASGGNMPLRKFSWLHSASGRVHVLEDRSLKLDVLTEDMGeyTCEADNAVGGITATGI 290
      : | | : | : : : : | : | : | | | | : |
Db    258 K--EGGELPANRTFFENF-----KTKLKIIDVSEADSGNYKCTARNTLGSTHHVIS 306

Qy    291 LTVHAPPKFVIRPKNLQVEIGDEVLFECQANGHPRPTLYWSVEGNSLLLPGYDGRMEV 350
      : | | | : : | : : : | : | | | : | : | : | : |
Db    307 VTVKAAPYITAPRNLVLSPGEDGTLICRANGNPKPSISWLTNGVPIAIP--EDPSRKV 364

Qy    351 TLTPEGRSVLSIARFAREDSGKVVTTCNALNAVGSVSSRTVVSVDTFELPPPIEQGPVN 410
      : | : : | : : | | | | : : : | : : | | : | |
Db    365 ---DGDITIFSA--VQERSSAVYQCNASNEYGYLLANAFVNLAE---PPRILT--PAN 413

Qy    411 QTLPV--KSIVLPCRTLGTVPVQVSWYLDGIPIDVQEHERRNLSAGALITISDLQRHED 468
      : | | | : : | : | : | : | : | | | | : : :
Db    414 KLYQVIADSPALIDCAYFGSPKPEIEWF-RGVKGSILRGNEYVFDHNGTLEIPVAQR-DS 471

Qy    469 EGLYTCVASNRNGKSSWSGRLYLDTPT---NPNKIFRA-----PEL---- 507
      | | | | | : | : | : | | : : : | : : | |
Db    472 TGYTTCVARNKLGKTQNEVQLEVKDPTMIKQPQYKVIQRSQAQSFECVIKHDPTLIPTV 531

Qy    508 -----STY----- 510
      : : : : :
Db    532 IWLKDNNELPDERFLVGKDNLTIMNVTDKDDGTTCIVNTILDSVSASAVLTVAAPPT 591

Qy    511 -----PGPPGKPMVEKGENSVTLSWTRSNKVGSSSLVGVIEM-FGKNETDGWVAVG 562
      | | : : : | : | | : : : | | | : | |
Db    592 PAIIYARNPNPLDLELTQLERSIELSWPGEE--NNSPITNFVIEYEDGLHEPGVWHYQT 650

Qy    563 TRVQNTTFTQTGLLPGVNYFFLIRAENSHGLSLPSPMSEPITVGTFRYNSGLDSEARAS 622
      : | : | | | | : | | | | | | : | : |
Db    651 EVPGSQTTVLQLSPYVNYSFVRVIAVNEIGRSQPSSE-----QYLTKSANPDE---- 700

Qy    623 LLSGDVVELSNASVVD--TSMKLTVIINGKYVEG-----FYVYARQLPNIVNPNAPVT 676
      | | : | : : | : : | : | | |
Db    701 -----NPSNVQIGSEPDNLVITWESLKGFGSQNGPGLQYKVSWRQKDVEDDEWISVVVA 753

Qy    677 SNTNPLLGSTSTSSASASASAL--ISTKPNIAAAGKRDE-----TN 717
      : : : : | | : : | : : | : : |
Db    754 NVSKYIVSGPTTFVPYIEIKVQALNDLGYAPEPESEVIGSHGEDLPMVAPGNVQVHVINSTL 813

Qy    718 QSGGGAPTPL-----NTKYRMLTLNKGGAASSCTITGLV 751
      | | | : : : | : : : | : : |
Db    814 AKVHWDPVPLKSVRGHLQGYKVYYKWQSLRRSRKRHVKKILTF--RGNKTFGLMPLGLE 871

Qy    752 QYTLTEFFIYVPFYKSVGEKPSNSRIARTLEDVPSEAPYMGAEALLNSSAVFLKWKAPELK 811
      | : | : : | | : : | | | | : : : | : | : |
Db    872 PYSSYKLNVRVYVNGKGEGPASPDKVFKTPEGVPS-PPSFLKITNPTLDSLTLWEGSP--T 928

Qy    812 DRHGVLNLYHVIRGIDTAHNFSRILTNVTIDAASPTLVLANLTGVMGY-----TVG- 863
      : | | : | : : | : | : | : | | : |
Db    929 HPNGVLTSTYLKFPINNTHELGP-LVEIRIPANESSLLNLNLYSTRYKFYNAQTSVGS 987

Qy    864 -----VAGGNAGVGYPYCPATLRDLPITKRL-----DPFIN-----QRDHVN 901

```

Db 988 GSQITEEAVTIMDEAGILRPAVGAG-KVQPLYPRIRNVTTAAETIANISWEYEGPDHAN 1046
Qy 902 ----- 901
Db 1047 FYVEYGVAGSKEDMKKEIVNGSRSPFFVLKGLTPGTAYKVRVGAELSGFRSSEDLFETGP 1106
Qy 902 -----DVLTOEWFIIILGAILAVLMLSPGAMVFKRHHMMKQALNTRMGNHSDVL 954
Db 1107 AMASRQVDIATQCFIQLMCAV-ALLILILLVCFIRN-----KGG----- 1147
Qy 955 KMPSLSARNGNGYVLDSTGGMVRPSPGGDSLEMQKDIADYAPVCGAPSPAGGTTSS 1014
Db 1148 KYPVKEKEDAHAF-DPEIQPMKEDDGTGFEYRSLESD-AEDHKPLKKSRTSPSDRTVKK 1203
Qy 1015 GSGGGA---GSGASGGDDIHGGHGSERNQRYVGEYSNIPTDYAEVSSFGAPSEYGRH 1070
Db 1204 EDSDDLVDYDGEYNGQFNEGDS-----FIGQYS-----GKKEKEPAE-GNE 1244
Qy 1071 GNASPAP 1077
Db 1245 SSEAPSP 1251

RESULT 14
S26180
neurofascin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jan-2000
C:Accession: S26180
R:Volkmer, H.; Hassel, P.; Wolff, J.M.; Frank, R.; Rathjen, F.G.
J. Cell Biol. 118, 149-161, 1992
A:Title: Structure of the axonal surface recognition molecule neurofascin and its relation to the cell adhesion molecule L1; fibronectin type III repeat homology; L1
A:Reference number: S26180; MUID:92317154
A:Accession: S26180
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1272 <VOL>
A:Cross-references: EMBL:X65224; NID:g63659; PIDN:CAA46330.1; PID:g63660
C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; L1
F:279-336/Domain: immunoglobulin homology <IMM>

Query Match 7.8%; Score 565.5; DB 2; Length 1272;
Best Local Similarity 21.1%; Pred. No. 1.2e-22;
Matches 293; Conservative 160; Mismatches 455; Indels 479; Gaps 55;

Qy 7 IEHPMDTIV-----PKNDPFTFNCQAEQNPPTTIQWFKDGR--- 42
Db 26 IEVPLDSNIQSEIQPPTITKQSVKDYIVDPNRNIF-IECEAKGNPVPTFSWTRNGKFFN 84
Qy 43 -----ELKTEGSHRIMLPAGGLFLLKVIHSRRESAGTYWCEAKNEFGVARSRNATL 95
Db 85 VAKDPKVMRRRSGTLVIDFHGG-----RPDDYEGEYQCFARNDYGTALSSKIH 135
Qy 96 QVAVLRDEFRLPAN---TRVAQGEVALMECGAPRGSPEQISWRKNGQTLNLVGNKRI 151
Db 136 QVS---RSPLWPKEKVDIEVDEGAPLSLQCNPPGLPPPVIFWMSSSME-PIHQDKRV 190
Qy 152 RIVDGG-----NLAIQEARQSDGGRYQCVK-----NVVGTRESATAFLK 191
Db 191 SQQGNDLYFSNVMLQDA-QTD---YSCNARFHTHTTIQKNPYTLKVKTKRPHNETSLR 246
Qy 192 VHVPRPFLIRG-----PQNQTAVGVSSVFCQRIIGDPLPDVLRRTASGGN 237
Db 247 NHTDMYSARGVTETTPSFMYPGTSSQMYLGRVDLLLECIASGVPAIDIMYK--KGE 304
Qy 238 MPLRKFSWLHSASGRVHVL---DRSLKLDVDTLEDMEYTCADNAVGGITATGILTVA 295
Db 305 LPAGKTK-----LENFNKALRISNVSEEDSGEYFCLASNMKGSIRHTISVRVKA 353
Qy 296 PPKFVIRPKNLQVEIGDEVLFECQANGHPRTPLYSVEGNSLLLPYGRDGRMEVTLTPE 355

Db 354 APYWLDEPQNLILAPGEDGRLVCRANGNPKPSTIQWLVNGEP-----IEGSPPNP 402
Qy 356 GRSVLS---IARFAREDSGKVVTCNALNAVGSVSSRTVSVSDTQFELPPPIIEQGPVNT 412
Db 403 SREVAGDTIVFRDTQIGSSAVYQCNASNEHGYLLANAFVSF---LDVPPRIL--APRNQL 457
Qy 413 LPV--KSIVVLPCTRLTGPVQVSWYLDGIPIDVQEHERRNLSDA-----GALTISD 462
Db 458 IKVIQYNRTRLDCPFFGSPITLRFWFKNG-----QGNMLDGGNYKAHENGSLMS- 507
Qy 463 LQRHEDEGLYTCVASNRNGKSSWSGYLRDLTPNPNKFFRAPE----- 506
Db 508 MARKEDQGIYTCVATNILLGKVEAQVRLEVKDPT----RIVRGPEQVVKRGSMPRLHCRV 563
Qy 507 -----LSTY----- 510
Db 564 KHDPTLKLTVTLWKDDAPLYIGNRMKKEDGLTIYVGAEKDQGDYTCVASTELDKDSAKA 623
Qy 511 -----PGPPGKPMVEKGENSVTLWSRNSKVGSSLVGVVEMFG 551
Db 624 YLTVLAI PANRLRDLPKERDRPDRLSLSLAERSVKLTWIPGDD--NNSPTIDYIVQFEE 682
Qy 552 KNETDG--WVAVGTRVQNTTFTQTGLLPVNYFFLIRAENSHGLSLPSMSEPTIVGTRYF 610
Db 683 DRFPQGTWHNHSRYPGNVNSALLSLSPYVNYQFRVIAVNDVGSLSMPSE-----RYQ 736
Qy 611 NSGLDLSEARASLLSGDVVELSNASVVDSTSMKLTWQIINGKYVEG---FYV--YARQLP 665
Db 737 TSG-----ARPEINPTGV---QAGATQKNMEITWPLNATQAYGNLRIVWRRRDP 787
Qy 666 ----NPIVNNPAPVTSNTNPLLGSTSTSSASASASALISTKPNIAAAGKRDETQSG 720
Db 788 RGSWYNETVKAPRHVVWNT-PIY-----VPEIKVQAENDFGRAPEPETIYIGYS 836
Qy 721 GGAPTPLNTKIRMLTILNG----- 739
Db 837 EDPYKAAPT DVR-IRVLNSTAIALTWTRVHLDTIQGLKEYRAYFWRDSSLKNLWVSKK 895
Qy 740 -----GGASSCTITGLVQYTYEFFIVPFYKSVGEKPSNSRIARTLEDVPSEAPYGM 792
Db 896 RQYVSFPGDRNRGIVSRFLFPYSNKLEMMVTVNGRGDGPSEVKEFPTPEGVSSSPRY-LR 954
Qy 793 ALLNSSAVFLKWKAPELKDRHGVLNHYHIVR-----GIDTAHNSRILNTVTIDAA 845
Db 955 IRQPNLESINLEWDHPE--HPNGVLGTYNLRYQAFNGSKTGRTLVNTPSPNQTRFTVQRT 1012
Qy 846 SPTLVLANLTGVMYTVGAAGNAGVGPYCV-----PATLRDLPIKRLDP 892
Db 1013 DPI-----SRYRFFLRARTQVGDGEVIVEESPALLNEATPTPASTWLPPTTELTP 1063
Qy 893 ----- 892
Db 1064 AATIAATTTTATPTTPTTPTTPTTPTTPTTATTTTATAASTVASTTTTAERAAAATKQE 1123
Qy 893 FINQRDHVNDVLTQPFIIILGAILAVLMLSPGAMVFKR---KEMMM--KQALNTRM 946
Db 1124 LAYTKNHV-DIATQGWFIQLMCAI-ALLVLILLVCFIKRSRGKYPVRDNKDEHLNPD 1181
Qy 947 GNHSDVLKMPSLSARNGNGYVLDSTGGMVRPSPGGD-----SLEMQK--DHIADYAP 999
Db 1182 KNVEDGSFDYRSLESDEN-----KPLPNSQTSLDGTIKQESDSDSLVDY-- 1226
Qy 1000 VCGAPGSPAGGTTSSGGSGAGSGASGGDDIHGGHGSERNQRYVGEYSNIPTDYAEVSS 1059
Db 1227 -----GEGGEGQFNEGDS-----FIGQYT-VKKOKEETE- 1254
Qy 1060 FGKAPSE 1066
Db 1255 -GNESSE 1250

RESULT 15
150600

Qy 10 PMDPTVPKNDPFTENCAEGNPTTIQWFKDGRELKTDGTSRIMLPAGGLFFLKVIHSR 69
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 25 PMDILSVRGASVYINCSSYCETPPKIEWKDKGTLLNLVSDDRQLLPDGSLLINSVVHSK 84

Qy 70 -RES DAGTYWCEANE-FGVARSRNATLQVAVLREDFLEPANTRVAQGEVALMECCAPR 127
: | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 85 HNKPDDEGYQCVAIVESLSGISVRTAKLTAVAGL-PRFSTSQPELSSVYKGN SAILNCEV-N 142

Qy 128 GSPEQISWRKNQOTLNLVGNKRIRIVDDGGNLAIQEARGSDDGRIYCVVKNVVTRESAT 187
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 143 VDLAPFVRWEQDRQPLSL--DDR VFKLPSGALLIGNATDTIDGGFYRCVIESGGTPKYSEE 200

Qy 188 AFLKVHVHRP-----FLIRGPQNQTAVVGSSVFQCRIGDDLDPVLWRRTASGGNMPLR 241
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 201 AELKILPDPEEPQSLVFVRQPSSLTKVTGQNAVPCVAGGFPTPYVRN--TKNGEEL--- 255

Qy 242 KFSWLHSASGRVHVLEDRLSKLDDVTLEDMEGYTEADNAVGGITATGILTVHAPPKFVI 301
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 256 ----ITEDSERFALRAGSLLISDVTEEDVGTTCIADNENETIEAQAE LAVQVPPEFLK 311

Qy 302 RPKNLVEIGDEVLFECQANGHPRTLYWSVEGNSLLLPYRDGRMEVTLTPEGRSVLS 361
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 312 RPANIYAHESMDIYFECEVTGKTPTPVKWVNKG D-VVIP----- 349

Qy 362 IARFAREDSGKVVTCNALNAVGSSSR TVVSVDTFELPPPPIEQGPVNQTL PVKSIVVL 421
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 350 -----SDYFKIVKEHNQLVGLVK S----- 369

Qy 422 PCRTLTPVPQVSWYLDGIPIDVQEHERRNLS DAGALTISDLQRHEDEGLYTCVASNRNG 481
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 370 -----DEGFYQCI AENDVG 383

Qy 482 KSSWSG---YLRLDT--PTNPNIKFRAPELSTYP---GP-PGKPMQVEK---GENSVTL 529
: | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 384 NAQAQAQIILLDLDAIPTLPPTSLSATNDHLAPATTGPLTAPRDVVATLVSTRIRL 443

Qy 530 SW-TRSNKVGSSSL---VG YVIEMFGKNETDGWVAVGTRVQNTT---FTQT---GLLP GV 579
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 444 TWRTVPSPDPOGDNLNTYSIFPTKEGINRE-----RVENTS RGETQVMIQNLMPET 493

Qy 580 NYFFLIRAENSGLSLPMSSEPI TVGTRYFN SGDL LSEARASLLSGDVVELSNASVSDS 639
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 494 VYVFRVAQNHKG---HGESSAPLKVATQ----- 519

Qy 640 TSMKLTWQJINGKYVEGFVYARQLPNP IVNNPAVTSNTWNPLL GSTS STSASASASAL 699
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 520 -----PEVLQPGA----- 528

Qy 700 ISTKPNI AAAGRKDGETNQSGGAPTPLNT KYRMILTILNGGG----- 741
| | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 529 ---PNIRAY-----AGSP TSVTVTWE---TPLSGNGEI QNYKLIYMKGQDQSE 571

Qy 742 ----ASSCTITGLVQYTLYEFFIVPFYKSV EKGPSNSRIARTLEDVPSEAPYGMEALL 796
| | | : | | | : | | | : | | | : | | | : | | | : | | | :

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:18 ; Search time 162.41 Seconds
(without alignments)
274.602 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTVPKNDP.....RSLLSNSGSGTSSQAGHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597.5	8.2	1493	1 NE01_MOUSE	P97798 mus musculus
2	589.5	8.1	1377	1 NE01_RAT	P97603 rattus norv
3	575.5	7.9	1461	1 NE01_HUMAN	Q92859 homo sapien
4	568	7.8	1036	1 AX01_CHICK	P28685 gallus gall
5	560	7.7	1443	1 NE01_CHICK	Q90610 gallus gall
6	555.5	7.6	1284	1 NRCA_CHICK	P35331 gallus gall
7	552.5	7.6	1447	1 DCC_MOUSE	P70211 mus musculus
8	551.5	7.6	1447	1 DCC_HUMAN	P43146 homo sapien
9	544.5	7.5	1040	1 AX01_HUMAN	Q02246 homo sapien
10	540.5	7.4	1040	1 AX01_RAT	P22063 rattus norv
11	538	7.4	2012	1 DSCA_HUMAN	O60469 homo sapien
12	536.5	7.4	1018	1 CONT_HUMAN	Q12860 homo sapien
13	534	7.3	1239	1 NRG_DROME	P20241 drosophila
14	531.5	7.3	1020	1 CONT_MOUSE	P12960 mus musculus
15	521	7.2	2029	1 LAR_DROME	P16621 drosophila
16	517.5	7.1	1010	1 CONT_CHICK	P14781 gallus gall
17	511.5	7.0	1260	1 CAML_MOUSE	P11627 mus musculus
18	509	7.0	1257	1 CAML_HUMAN	P32004 homo sapien
19	497.5	6.8	1259	1 CAML_RAT	Q05695 rattus norv
20	495.5	6.8	1897	1 PTFP_HUMAN	P10586 homo sapien
21	487	6.7	1266	1 NGCA_CHICK	Q03696 gallus gall
22	485.5	6.7	1912	1 PTPD_HUMAN	P23468 homo sapien
23	457.5	6.3	1913	1 KMLS_HUMAN	Q15746 homo sapien
24	441	6.1	3707	1 PGBM_MOUSE	Q05793 mus musculus
25	427.5	5.9	4393	1 PGBM_HUMAN	P98160 homo sapien
26	412.5	5.7	1070	1 PTK7_HUMAN	Q13308 homo sapien
27	406	5.6	1091	1 NCA1_CHICK	P13590 gallus gall
28	403.5	5.5	761	1 NCA2_HUMAN	P13592 homo sapien
29	398.5	5.5	837	1 NCM2_HUMAN	O15394 homo sapien
30	396.5	5.5	853	1 NCA1_BOVIN	P31836 bos taurus
31	389.5	5.4	837	1 NCM2_MOUSE	O35136 mus musculus
32	389.5	5.4	1115	1 NCA1_MOUSE	P13595 mus musculus
33	389	5.3	1088	1 NCA1_XENLA	P16170 xenopus lae

34	388.5	5.3	725	1 NCA2_MOUSE	P13594 mus musculus
35	388.5	5.3	848	1 NCA1_HUMAN	P13591 homo sapien
36	388.5	5.3	1092	1 NCA2_XENLA	P36335 xenopus lae
37	385.5	5.3	858	1 NCA1_RAT	P13596 rattus norv
38	380	5.2	1906	1 KMLS_CHICK	P11799 gallus gall
39	368	5.1	1051	1 PTK7_CHICK	Q91048 gallus gall
40	356	4.9	1666	1 SKLM_MOUSE	Q62234 mus musculus
41	355.5	4.9	2481	1 UN52_CAEEL	Q06561 caenorhabdi
42	346	4.8	811	1 FS22_DROME	P34083 drosophila
43	339	4.7	873	1 FS21_DROME	P34082 drosophila
44	323	4.4	1451	1 MYM1_HUMAN	P52179 homo sapien
45	321	4.4	898	1 FAS2_SCHAM	P22648 schistocerc

ALIGNMENTS

RESULT 1
NE01_MOUSE
ID NE01_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NE01 OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED UBIQUITOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y09535; CAA70727.1; -.
DR HSP; P02751; ITTG.
DR MGD; MGI:1097159; NE01.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.

[illegible]

```

RESULT 3
NEO1_HUMAN
ID NEO1_HUMAN STANDARD; PRT; 1461 AA.
AC Q92859; O00340;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEO1 OR NGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE=97236653; PubMed=9121761;
RA Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT "Identification and characterization of neogenin, a DCC-related
RT gene.";
RL Oncogene 14:1129-1136(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE=97312699; PubMed=9169140;
RA Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA Korenberg J.R.; Dreyer W.J.;
RT "Molecular characterization of human neogenin, a DCC-related protein,
RT and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT q23.";
RL Genomics 41:414-421(1997).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC LINES.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X63101; CAA44815.1; -.
DR PIR; S22128; S22128.
DR PIR; S22383; S22383.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
PFAM; PF00041; fn3; 3.
PFAM; PF00047; ig; 6.

Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
Cell adhesion; Repeat.

FT SIGNAL	1	23	OR 25 (POTENTIAL).
FT CHAIN	24	1036	AXONIN-1.
FT PROPEP	?	1036	REMOVED IN MATURE FORM.
FT DOMAIN	49	113	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	143	211	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	249	308	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	336	397	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	428	490	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	518	589	IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN	599	608	HINGE (POTENTIAL).
FT DOMAIN	601	607	GLY/PRO-RICH.
FT DOMAIN	608	709	FIBRONECTIN TYPE-III.
FT DOMAIN	710	811	FIBRONECTIN TYPE-III.
FT DOMAIN	812	912	FIBRONECTIN TYPE-III.
FT DOMAIN	913	1009	FIBRONECTIN TYPE-III.
FT MOD_RES	724	724	BLOCKED.
FT CARBOHYD	71	71	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	199	199	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	456	456	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	472	472	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	493	493	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	770	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	900	900	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	914	914	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE	1036 AA;	113301 MW;	08B80143BE779794 CRC64;

Query Match 7.8%; Score 568; DB 1; Length 1036;
Best Local Similarity 23.1%; Pred. No. 2.9e-23;
Matches 250; Conservative 116; Mismatches 352; Indels 364; Gaps 40;

Qy	4	PRIIEHPMDTTPVK---NDPFTFNCQAEQNPPTTIQWFKDRELKTDGTGSHRIMLPAGGL	60
Db	32	PVFEEQPAHTLFPESGAEEKVTLTCRANPPATYRWKNGTELMGPDGSRYLVDGL	90
Qy	61	FFLKVIHSRRSDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFRLE-----	107
Db	91	VISNPVKAK---DAGSYQCVATNARGTVVSREASLRFLQ-EFSAEERDPVKITEGWG	146
Qy	108	-----	107
Db	147	MFTCSPPPHYPALSYRWLLNEFPNIPADGRRFVSQTTGNLYIAKTEASDLGNYSFATS	206
Qy	108	-----PANTVAQGEVALMECGAPRGSPPEPQ	133
Db	207	HIDFITKSVFSKSQLSLAEDARQYAPSIKAKPPADTYALTQGMVLTLCFA-FGNVPVQ	265
Qy	134	ISWRK--NGQTLNLVGNKRIRIVDGGNLAIQEARQSDGRIQCVVKNVGTRESATAFLK	191
Db	266	IKWRKLDGSGTSKWLSEPL-----LHQNVDPEDEGTIECAENIKG-RDIYQGRII	317

Qy	192	VHVRPFLIRGPQNTAVVGSSVVFQCRIGGDPLDVLWRRRTASGGNMLPKFVSWLHSASG	251
Db	318	IHAQPDWLDVITDTEADIGSDLRWSCVAGSKPRPAVRWLR---DQQL-----ASQN	366
Qy	252	RHVLEDRLSLKDDVTLEDMEYTCADNAVGGITATGILTVHA-PPKFVIRPKNLVLEI	310
Db	367	RIEV-SGGLRFSKLVLEDGSMYQCAENKKGTVYASAEITVQALAPDFRLNPVKRLIPA	425
Qy	311	--GDEVLFECQANGHPRTLYWSVEGNSLLPGYDRGRMEVLTPEGRSVLSIARFARE	368
Db	426	ARSGKVIIPCPRAAPKATVLTW---KGTELLTNSSR---VTITADGTLILQ--NISK	476
Qy	369	DSGKVVTCNALNAVGSVSSRTVSVSDTQFELPPPIEBQGPVNTLPVKSIVVLPCTLTG	428
Db	477	DEGK-YTCFAENFMGKANSTGILSVRDATAK-----ITLAPSSADINVGENITLQCHASHD	530
Qy	429	PVPQV--SWYLDGIPIDVQEHE---RRNLSDA-GALTISDLQ-RHEDEGLYTCVASNRN	480
Db	531	PTMDLTFTWSLDDFPIDLDKSEGHYRRASVKEAVGDIAIVNAQLKH--SGRYTCTAQTVV	588
Qy	481	GKSSWSGYSRLDTPNPNKIFRAPELSTYGPFGPKQPMVEKGENSVTSLWTRSNKVGGS	540
Db	589	DSTSESATLTVRGP-----PGPPGVVVRDIGDTTVQLSWSRGFD-NHS	631
Qy	541	SLVGVIIE---MFGKNETDGVAVGTRVQN---TFTQTGLLPGVNYFFLIRAENSH	591
Db	632	PIARYSIEARTLLSNK-----WKQMRTPNVNIEGNAETAQVYNLIPWMDYEFVLASNIL	686
Qy	592	GLSLPSPMSEPIITVGTTRYFNSGLDSEARASLLSGDVVELSNASVVDSTSMKLTQIING	651
Db	687	GVGEPS-----	692
Qy	652	KYVEGFVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSASASASALISTK---PNIAA	708
Db	693	-----LP-----SSKIRTKEAAPTVPAP	709
Qy	709	AGKRDGETNQSGGGAP-----TPLNTKYRMLTILNGGG-----ASSCTITGLV	751
Db	710	SGL-----GSGGGAPNELIINWPTLRDYQ---NGDGFYILSFRKKGTOGWLTVARV	758
Qy	752	-----QYLYEFFIVFPYKSVGEKPSNSRIARTLEDVPSEAPYGMALL	795
Db	759	PHAESLHYVYRNESIGPYTPFEVKIKAYNRKGEPSLTAIVYSAEPEKVPFRTAKA	818
Qy	796	LNSSAVFLKWKAPKELDRHGVLLNHYHV-IVRGIDTAHNFSRLTNVTIDAASPTLVLANL	854
Db	819	VLSSEMDVSWEPEVQDGTGVLGVEIRYWKDGEKAAADRVTAGLVTSAHVT---GL	874
Qy	855	TEGVMYTVGAAGNAGVGPYC-----VPATLRLDPITKRLDPFINQ	896
Db	875	NPNTKYHVSVRAYNRAGAGPPSPSTNITTKPPRRPPGNISWTLTGSTVTIKWDPVVAQ	934
Qy	897	RD 898	
Db	935	AD 936	

RESULT 5

NEOL_CHICK

ID NEOL_CHICK STANDARD; PRT; 1443 AA.

AC Q90610;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NEOGENIN (FRAGMENT).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;

RX MEDLINE=95105243; PubMed=7806578;

```

RESULT      6
NRCA_CHICK
ID  NRCA_CHICK      STANDARD;      PRT;  1284 AA.
AC  P35331;
DT  01-FEB-1994 (Rel. 28, Created)
DT  01-FEB-1994 (Rel. 28, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  NG-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RC  STRAIN=WHITE LEGHORN; TISSUE=EMBRYONIC BRAIN;
RX  MEDLINE-91258407; PubMed=2045418;

```

```
Qy      8 EHPMDTTVPKNDDPTFNCAEQAGNPTPTIQFQKDGRELKDTGSHRIMLPAGGLFLKLVKH 673  
       : | | : | : | : | : | : | : | : | : | : | : | :  
Db     46 QSPKDYIVDPRENVIQCEAKGPKPPSFSWTRNGTHFDIDKDAQVTMKPNSGTLVVNMN 105  
  
Qy     68 S-RRESDACTYWEAKNEFGVARSRNATLQV--AVLRDEFRLPANTRVAQEVALMECG 124  
       : | : | | : | : | : | : | : | : | : | : | : |  
Db    106 GVKAEAYEGVYQCTARNERGEAAISNNVIRPSRPLWTKEKLEPNHVR--EGDSLVLNCR 163  
  
Qy    125 APRGSPEPQISWRKNQGTLNLVGNKRIIRDGGNLAIQEARQSDDDR--YQCJV----- 176  
       | | | | | | | | : | : | : | : | : | : | : |  
Db    164 PPVGLPPPFIIFWMDNA-FQLRLQPSERVSQGLNGDLIYFSNV-QPEDTRVDYICYARFNHTQ 221  
  
Qy    177 ---KNVGTRESATAFLKVHRVPRFLIRGP---QNQTAVVGSSVVFQCRIGGDPLPDVL 228  
       | : : : | | | | : | : | : | : | : | : | : |  
Db    222 TIQQKPISVXFSTK--PVTERPPVLLTPMGSTSNNKVELRGVNLLLECIAAGLPTPVIR 279  
  
Qy    229 WRRITASGGNMLRKFSWLHSASGRVHVLEDRLSKLDVLTLEMGEYTCEADNAVGGITAT 288  
       | : | : | : : : : | : | : | : | : | : | : |  
Db    280 WIK--EGGELSPANRTFFENF-----KTKLIIDVSEADSGNFKCTARNTLGSTHHV 328  
  
Qy    289 GILTVHAPKFVIRPKNLVEIGDEVLFECQANGHPRPTLYWSVEGNSSLLPGYRDGRM 348  
       : | | : | : : : : : | : | : | : | : | : | : |  
Db    329 ISVTVKAAPPYITAPRNVLVSPGEDGTLCIRANGNPKPSISWLTVGNPIAIAP--EDPSR 386  
  
Qy    349 EVTLTPEGRSVLSIARFAREDSGKVVTCTNALNAVGSVSRTTVSVSDTFELPPPIIEQGP 408  
       : | : | : | : | : | : | : | : | : | : | : | : |  
Db    387 KV---DGTIIIFA--VQERSSAVYQCNASNEYGYLLANAFVNLAE---PPRILT--P 435  
  
Qy    409 VNQTLPV--KSIIVLPCRTLGTVPVQSWYLDGIPIDVQEHERRNLSAGALITISDLQRH 466  
       | : | | : | : | : | : | : | : | : | : | : | : |  
Db    436 ANKLQVIADSPALIDCAYFGSPKPEIEWF-RGVKGSILRGNEYVHDNGTLEIPVAQK- 493  
  
Qy    467 EDEGLTYCVASNRRNGKSSWSGYLRDLTPT---NPNIKFFRA-----PEL-- 507  
       : | | | | : | : | : | : | : | : | : | : | : |  
Db    494 DSTGYTTYCARNKLKTQNEVQLEVKDPTMIKQPQYKVIQRSQAQSFECVHKDPTLIP 553  
  
Qy    508 -----STY----- 510  
           ||  
Db    554 TVIWLNKDNEGPDDERFLVGKDNLTIMNVTDKDDGYTCIVNTTLDVSASAVLTVVAAP 613  
  
Qy    511 -----PGPPGQMVEKGENSEVTLSWTRSNKVGSSSLVGTVIEM-FGKNETDGWA 560  
       : | | : : : | : | : | : | : | : | : | : | : |  
Db    614 PTPAIIYARPNPDLDELTLQELERSIELSWVPGE--NNSPITNFVIEYEDGLEHPGVW-H 671  
  
Qy    561 VGTRVQNT-TFTQTGLLPGVNYFFLIRAENSAGLSLSPMSEPITVGTRYFNGLDLSEA 619  
       | | : | | | | : | | | | | : | : | : | : |  
Db    672 YQTEVPGSHTTVQLKLSFYVYSFRYIAVNEIGRSQSPSEPSE-----QYLTKSANPDE- 724  
  
Qy    620 RASLLSGDVVLESNASVDS--TSMKLTWOIINGKYVEG---FYVYAROINPNIIVNNA 673
```

Db 725 -----NPSNVQGIGSEPDNLVITWESLKGQSQNGPGLQYKVSWRQRDVEDWTSV 774

Qy 674 PVTSTNTNPLLGSTSTASASASASAL--ISTKPNIAAAGKRDGE----- 715
| : : : | | : : : | : : : |

Db 775 VVANVSKYIVSGCTPTFPYPIKVALNDLGYAPESEVIGHSGEDLPMVAPGNVQVHVIN 834

Qy 716 -TNQSGGGAPTPL-----NTKYRMLTILNGGGASSCTIT 748
| | | | : : : | : : : |

Db 835 STLAKVHWDVPVPLKSVRGHLQYKVVYWKVQSLRRSKRRVVEKKILTF--RGNKTFGMPL 892

Qy 749 GLVQYITLYEFFIVPFYKSVGEKPSNSRIARTLEDVPSAPYGMALLNSSAVFLKWKAP 808
| | : : : | | : : : | | : : : | : : : |

Db 893 GLEPYSSYKLVNRVVNGKEGPASPDVKFETPEGVPS-PPSFLKITNPILDSLTLEWGSP 951

Qy 809 ELKDRHGVLNTHVIVRGIDTAHNFSRIILTNTVIDAASPTLVLANLTGVMTYGVGAAGN 868
: | | : : : | : : : | : : : | : : : |

Db 952 --THPNGLTSYILKFQPINNTHELGP-LVEIRIPANESSLILKLNLYSTRYKFYFNAQT 1008

Qy 869 NAGVGPPYCPVATL-----RLDPITKRL-----DPFIN-----QRD 898
: | | : : : | : : : | : : : |

Db 1009 SVSGSGSQITEAVTIMDEAGILRPVAGACKVQPLYPRINRVTTAAAEITYANISWEYEGPD 1068

Qy 899 HVN----- 901
| |

Db 1069 HANFYVEYGVGASCKEDWKKEIVNGSRSSFVLKGLTPGTAYKVRVGAEGLSGFRSSDDLFE 1128

Qy 902 -----DVLTPQWFIIILGAILAVLMLSPGAMVFVKRHHMMMQSALNTRMGNHTS 951
| : : : | : : : | : : : | : : : |

Db 1129 TGPAMASRQVDIATQGWFIGLMCAV-ALLILILLIVCFIRRN-----KGG--- 1172

Qy 952 DVLKMPSLSARNGNGYWLDSSTGGMVWRPSPGGDSLEMQKDHIADYAPVCGAGPSPAGGG 1011
| | : : : | : : : | : : : | : : : |

Db 1173 ---KYPVKEKEDAHA---DPEIQPMKEDDGTFGEYRSLESD-AEDHPLKKGSRTPSDRT 1225

Qy 1012 TSSGGSGGA---GSGASGGDDIHGGHGSERNQRRYVGEYSNIPTDYAEVSSFGKAPSEY 1067
| : | : | : | : : : | : : : |

Db 1226 VKKEDSDSLVDYGEVGNQFNEDGS-----FIGQYS-----GKKEKEPAE- 1266

Qy 1068 GRHGNASAP 1077
| : : : |

Db 1267 GNSESEAPSP 1276

RESULT : 7

DCC_MOUSE

```

ID DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RX MEDLINE=96112625; PubMed=8570174;
RA Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
RT gene (mDCC) and its expression in the developing mouse embryo.";
RL Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD

```

```

CC      FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
CC      ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC      DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC      BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC      -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC      EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC      AND REMAIN AT THIS LEVEL IN THE ADULT.
CC      -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X85788; CAA59786.1; -.
DR      HSSP; P56276; 1TLK.
DR      MGD; MGI:94869; DCC.
DR      INTERPRO; IPR001777; -.
DR      INTERPRO; IPR003006; -.
DR      PFAM; PF00041; fn3; 6.
DR      PFAM; PF00047; ig; 4.
DR      PRINTS; PR00014; FNTYPEIII.
KW      Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW      Anti-oncogene; Alternative initiation; Alternative splicing.
FT      SIGNAL      1      25      POTENTIAL.
FT      CHAIN       26      1447    TUMOR SUPPRESSOR PROTEIN DCC, LONG
FT                                     ISOFORM.
FT      CHAIN       85      1447    TUMOR SUPPRESSOR PROTEIN DCC, SHORT
FT                                     ISOFORM.
FT      INIT_MET    85      85      FOR SHORT ISOFORM.
FT      DOMAIN      26      1097    EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM    1098     1122    POTENTIAL.
FT      DOMAIN      1123     1447    CYTOPLASMIC (POTENTIAL).
FT      DOMAIN      54      124     IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      154     219     IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      254     317     IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      345     407     IG-LIKE C2-TYPE DOMAIN.
FT      DOMAIN      426     522     FIBRONECTIN TYPE-III.
FT      DOMAIN      525     618     FIBRONECTIN TYPE-III.
FT      DOMAIN      619     716     FIBRONECTIN TYPE-III.
FT      DOMAIN      722     816     FIBRONECTIN TYPE-III.
FT      DOMAIN      840     940     FIBRONECTIN TYPE-III.
FT      DOMAIN      941     1042    FIBRONECTIN TYPE-III.
FT      DISULFID    61      117     BY SIMILARITY.
FT      DISULFID    161     212     BY SIMILARITY.
FT      DISULFID    261     310     BY SIMILARITY.
FT      DISULFID    352     400     BY SIMILARITY.
FT      CARBOHYD    60      60      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    94      94      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    299     299     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    318     318     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    478     478     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    628     628     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD    702     702     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      VARSPLEC    819     838     MISSING (IN EMBRYONIC ISOFORM).
SO      SEQUENCE    1447; AA: 158298 MW: 0D1F1097C22D5B9F CRC64:

```

Query Match 7.6%; Score 552.5; DB 1; Length 1447;
Best Local Similarity 22.0%; Pred. No. 3.1e-22;
Matches 316; Conservative 172; Mismatches 532; Indels 415; Gaps 66;

Qy 7 IEHPMDTTPVKNDPFTFNCAQEGN-PTPTIQWFKDGRGLKTDGTGSHRIMLPAGGLFFLKV 65
: | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db 43 VSEPSDAVTFMGGNVLLNCSAESDRGVPVYIKWKDGLLALGMDDRKQLPNGLLIQNI 102
Qy 66 IHSR-RESDAQTYWCEAK-NEFGVARSRNATLQVA-VLRDEFLEPANTRVAQGEVALME 122
: | | : | | : | | : | | : | | : | | : | | : | | : | | :

```

Db      103 LHSRHHKPEDEGLYCEASLADSGSIISRTAKTVVAGPLR--FLSQTESITAFMGDTVLKK 160
Qy      123 CGAPRGSPPEPQISWRKNQGTLN-LVGNNKRIRIVDGGNLAIQEARGSDDGRIQCVCVKNVVG 181
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      161 CEV-IGEMPMTIHWQKNQDLNPLPGDSRVVVLPSGALQISRLQPGDGSVYRCARPAS 219

Qy      182 TRESATAFLKV-----HVRPFILIRGPQNQTAVVGSSVVFCRIGGDPLPDVLRWRTASG 235
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      220 IRTGNEAEVRILSDPGLHRQLYLQRPSNVIALTEGKDAVELECCSVGYPPPSFTWLARGEV 279

Qy      236 GNMPLRRKFSWLHASGRVHVLEDRLSKLDDVTLEDMEGYTCEADNAVGGITATGILTVAH 295
       :   |:|   :   |   :   |   :   |   |   |   |   |   |   |   |
Db      280 IQLRSKKYS-----LLGGSNLLISNVTDDSGTYTCVVITYKENENISAEBLTVLV 329

Qy      296 PPKFVIRPKNQLVIEIGDEVLFECQANGHPRTLYWSVEGNSLLLPGYRDGRMEVTLTPE 355
       ||   |   |   |   :   ||   |   |   ||   |   |   :   :||   :
Db      330 PPWFLNHPSNLYAYESMDIEFECAVSGKPVPTVNMWKNGD--VVIP---SDYPQIV---- 380

y      356 GRVSLSIARFAREDSGKVVTCTNALNAVGVSSSRTVVSVDTFELPPPPIEQGPVNQTLPV 415
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      381 GGSNLRILGVVKSDGE-FYCVAENEAGNAQS-----SAQLIVPKPAI---PSSSILPS 430

Qy      416 KSIVVLPCLRTLGTPVPQVSWYLDGIPIDV-----QEHERR-----NLSDAGA-- 457
       ||   |   :   :   :||   |   :   :   |   :   |   |   |   |   |
Db      431 APRDVLVP-LVSSRFVRLSW--RPPAEAKGNIQTTVFPSREGDNRRERALNTQPGLSQ 486

Qy      458 LTISDLQRHEDEGLYT--CVASNRNKGSSWSGYLRDLTPTNPNIKFFRAPELSTYGPFG 515
       ||:   :|   :   |:|   ||   |   :   :   ||   ||   ||   ||
Db      487 LTVGNL---KEAMITFRVVAYNEWGPGE-----SSQPIKAVTQPELVQ-PGPVB 532

Qy      516 KPQMVEKGENSEVTLTSWRNSKVGSSLVGYVIEMF-----GKNETDGMWAVGTVRQNTT 569
       |   |   :   :|   |   :   |   |   :   |   |   |   |   :
Db      533 NLHAVSTSPTSILITWEPPAYANG-PVQGY--RLCFTEVSTGKEQ-----NIEVDGLS 582

Qy      570 FTQTGLLPQVNVFFLIRAENSGLSLPMSPEPIVTGTRIFNSGLDSEARASLLSGDVV 629
       :   ||   |   |   |   |   |   |   :   |||   |   |   :   |   :|
Db      583 YKLEGLKFTEYTLRFLAYNRYG--PGVSTDITVVTL-----SDVPSAPPQNIS--- 630

Qy      630 ELSNASVVDSTSMKLTW-----QIIINGKYVEGFYV-----NQSOGG 659
       ||:|   |:|:|   ||   |   :   |   :
Db      631 ---LEVNSRSKVSWLPSPSGTQNG-FITGYKIRHKTTTRRGEMETLEPNNLWYLTG 685

Qy      660 -----YARQLPNPIVNNAP----VTSNT--NPL-----LGSTSTASASASASALIST- 702
       |:   |:   ||   |   |   |   |   |   :   :|   :   :   |:|
Db      686 LEKGSQSYSFQVSAMTVNGTGPPSNWYTAETPENDLDESQVPDPSSLHVRPQNCIMSW 745

y      703 ---KPNIAAAGKRDG-----ET-----NQSGGG 722
       |||   |   |   |   |   ||   |   |   |   |   |   |   |
Db      746 TPPLNPNIIVRGYIIGYCGVGSPYAETVRVDSKORYYSIERLESSSHYVISLKAFPNAEG 805

Qy      723 AP----- 724
       |
Db      806 VPLYESATTRSITDPTDPVDYTPLLDDFPSTSGPDVSTPMLPVGVQVAVALTHEAVRWSWA 865

Qy      725 ---TPLNTK---YRLMTI--LNGGGAS-----SCTITGLVQYTYLIEFFIVPFYK 765
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      866 DNSVPKNQKTSQDRVLYTRWRITSFSASAKYKSEDTSLSYSTATGLKPNMTIEFSVMYTKN 925

Qy      766 SVEGKPSNSRIARTLEDVPSEAPYGMEALLNSS--AVFLKWKAPELKDRHGVLNYHVI 823
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      926 RRSSTWSMTHAHATTYEAAPTSAKDLTVTITREGKPRAVISVQQPP--LEANGKITAYIL- 982

Qy      824 VRGIDTAHNFSRILTNV-----TIDAASPTLVLANLTGVMYTVGVAAGNAGVGPY 875
       |   |   :   |   |   |   |   |   |   |   |   |   |   |   |
Db      983 -----FYTLDKNIPIDDWMETISGDRLTHQIMDSLDTMYFYRIQARNVKGVGPL 1033

Qy      876 CVP---ATLRLDPITKR-----LDPFINQRDHVND----- 902
       |   |   |:|   |   |   |   |   |   |   |   |   |   |
Db      1034 SDPILFRTLKVEHPDKMANDQGRHGDDGYWPVDTNLIIDRSTLNEPPIGQMHPHGSVTPQ 1093

Qy      903 -----VLTPQWFIILLGAILAVLMLSFGAMVFKRHMMMKQASALNTMRGNHTSDVLKM 956
       |:   |   :   |:|   |:|   |   |   |   |   :   :   |   |
Db      1094 KNSNLVIT---VTVG-VLTVLVVVIYAVICTRRSSAOORRK-----RATHSG----- 1138

```

```
Qy      957 PLSLARGN-----GYWLDSTSGGM--VWRPS---PGGDSLEMKQDHIADYAPVC GAP   1004
          | | | :       | :         | : | | : | | :
Db     1139 ---SKRKGSQLRPDLWIHHEEMKNIEKPTGTDPAGRDSP IQS---CQDLTPVSHSQ   1193
          | : | | | | | : | : | : | : | : | :
Qy     1005 GSPAGGGTSSGGSG---GAGSGASGGDDIHGGHGSENRQQRYVG EYSNIPTDYAEVSSF   1060
          | : | | | | | : | : | : | : | : | :
Db     1194 SETQMGSKSASHSGQDTEDAGSSMSTLERSLAARRATRAKL MIPMEAQS---SNPAVVSAI   1251
          | : | : | : | : | : | : | : | :
Qy     1061 GKAPSEYGRHGNASPAPYATSSILSPHQOQQQQQPRYQQRVPV-----GYGLQRFMH   1112
          | : | : | : | | | : | | | | : | |
Db     1252 PVPTLESAYQPGILSP---TCGYPH-----PQFTLRPVFPPTLSVDRGFGAGR---   1297
          | : | : | : | : | : | : | : | :
Qy     1113 PHYQQQQHQQQQAQQTHQQHQALQQHQQLPPSNIYQQMSTTSEIYP TINTGPSRSV   1167
          | : | : | : | | | : | : | : | :
Db     1298 -----TOSYSEGPTTOOQPLPPA---OPEHPSESSEAPSRTIPTACV   1336
```

RESULT 8

DCC HUMAN

```

ID DCC_HUMAN STANDARD; PRT; 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011532; PubMed=7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RT tumorigenesis.";
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE=90100559; PubMed=2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers.";
RL Science 247:49-56(1990).
RN [3]
RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBELD EXONS).
RX MEDLINE=91121517; PubMed=1991322;
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA Oliner J.D., Kinzler K.W., Vogelstein B.;
RT "Scrambled exons.";
RL Cell 64:607-613(1991).
RN [4]
RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX MEDLINE=94245241; PubMed=8188295;
RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT "The DCC gene: structural analysis and mutations in colorectal
RT carcinomas.";
RL Genomics 19:525-531(1994).
RN [5]
RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX MEDLINE=94243823; PubMed=8187090;
RA Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT human esophageal squamous cell carcinomas and their relation to
RT metastasis.";
RL Cancer Res. 54:3007-3010(1994).
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL

```

CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC -1- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X76132; CAA53735.1; -
DR EMBL; M32292; AAA35751.1; -
DR EMBL; M32286; AAA52174.1; -
DR EMBL; M32288; AAA52175.1; ALT_SEQ.

DR EMBL; M32290; AAA52176.1; -
DR EMBL; M63696; AAA52177.1; -
DR EMBL; M63700; AAA52178.1; -
DR EMBL; M63702; AAA52179.1; -
DR EMBL; M63718; AAA52180.1; -
DR EMBL; M63698; AAA52181.1; -

DR PIR; A54100; A54100.

DR PIR; A40098; A40098.

DR PIR; A38442; A38442.

DR HSSP; P56276; 1TLK.

DR MIM; 120470; -.

DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00041; fn3; 6.

DR PFAM; PF00047; ig; 4.

DR PRINTS; PR00014; FNTYPEIII.

KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;

KW Anti-oncogene; Disease mutation; Polymorphism.

FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	1447	TUMOR SUPPRESSOR PROTEIN DCC.
FT	DOMAIN	26	1097	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1098	1122	POTENTIAL.
FT	DOMAIN	1123	1447	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	54	124	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	154	219	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	254	317	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	345	407	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	426	522	FIBRONECTIN TYPE-III.
FT	DOMAIN	525	618	FIBRONECTIN TYPE-III.
FT	DOMAIN	619	716	FIBRONECTIN TYPE-III.
FT	DOMAIN	722	816	FIBRONECTIN TYPE-III.
FT	DOMAIN	840	940	FIBRONECTIN TYPE-III.
FT	DOMAIN	941	1042	FIBRONECTIN TYPE-III.
FT	DISULFID	61	117	BY SIMILARITY.
FT	DISULFID	161	212	BY SIMILARITY.
FT	DISULFID	261	310	BY SIMILARITY.
FT	DISULFID	352	400	BY SIMILARITY.
FT	CARBOHYD	94	94	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	318	318	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	702	702	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	168	168	M -> T (IN OESOPHAGEAL CARCINOMA).
FT				/FTid=VAR_003909.
FT	VARIANT	201	201	R -> G.
FT				/FTid=VAR_003910.
FT	VARIANT	1375	1375	P -> H (IN A COLORECTAL CARCINOMA).
FT				/FTid=VAR_003911.
FT	CONFLICT	138	138	MISSING (IN REF. 3).
FT	CONFLICT	233	329	MISSING (IN REF. 3).
FT	CONFLICT	421	421	MISSING (IN REF. 3).

SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;

Query Match 7.6%; Score 551.5; DB 1; Length 1447;
Best Local Similarity 21.8%; Pred. No. 3.5e-22;
Matches 312; Conservative 177; Mismatches 542; Indels 397; Gaps 64;

Qy	5	RIIEHPMDTTPVKNPDPFTFNCQAEQN-PTPTIQWFKDGRLEKTDGSHRIMLPAGGLFFL	63
Db	41	RFLSEPSDAVTMRGGNVLLDCAESDRGVPVKKWKDGIHALGMDERKQSLNGSLLIQ	100
Qy	64	KVIHSR-RESDACTYWEAK-NEFGVARSRNATLQVA-VLRDEFRLPANTRVAQGEVAL	120
Db	101	NILSRHKKPDEGLYQCEASLGDSGSIISRTAKVAVAGPLR--FLSQTESVTAFMGDTVL	158
Qy	121	MECAGPRGSPPEQISWRKNGQTLN-LVGNKRIRIVDGGNLAIEARQSDGGRYQCVVKNV	179
Db	159	LKCEV-IGEPMTIHWQKNQDLTPGDSRVVLPSPGALQISRLQPGDIGIRCSARNP	217
Qy	180	VGTRESATAFLKV-----HVRPFLIRGPQNTAVVGSSVVFQCRIGGDLPLDVLWRRTA	233
Db	218	ASSRTGNEAEVRLSDPGLRHQLYFLQRPNSVVAIEGKDAVLECCVSGYPPPSFTWLGR	277
Qy	234	SGGNMPLRKFSSWLHSASGRVHLEDRSLKDDVLEDMEGYTCEADNAVGGITATGILTV	293
Db	278	EVIQLRSKKYS-----LLGGSNLLISNVTDDSGMYTCVVITYKNENISASAEITV	327
Qy	294	HAPPKFVIRPKNLVEIGDEVLFECQANGHPRTLYSVEGNSLLLPYDGRMEVTLT	353
Db	328	LVPWFNLNHPNLYAYESMDIEFECTVSGRPVPTVNMKNGD--VVIP--SDYFQIV--	380
Qy	354	PEGRSVLSIARFAREDSGKVVTCNALNAVGSVSSRTVVSVDTFELPPPIIEQGPVNQTL	413
Db	381	--GGSNLRILCVVKSDEG-FYQCAENEAQNAQT-----SAQLIVPKPAIPSSSVLPSA	431
Qy	414	PVKSIVVLPCRTLGTVPVQVSWYLDGIPIDV-----QEHER--NLSAGA	457
Db	432	PRDVVPLV---VSSRFVRLS---RPPAEAKGNIQFTVFFSREGDNRERLNTQPGS	484
Qy	458	--LTISDLQRHEDEGLYT--CVASNRNGKSSWGYLRDLTPNPNIKFFRAPELSTYGP	513
Db	485	LQLTVGNL--KPEAMTYFRVVAINEWGPGE-----SSQPIKVAQPELV-PGP	530
Qy	514	PGKPMVEKGENSVTLSTWRSNKVGGSSLVGYVIEMF-----GKNETDGVAVGTRVQN	567
Db	531	VENLQAVSTSPSTILITWEPAYANG-PVQGY--RLFCTEVSTGKEQ-----NIEVDG	580
Qy	568	TFTQTGLLPGVNYFFLIRAENSHGLSLSPMSPEITVGTFRYNSGLDLSEARSLLSGD	627
Db	581	LSYKLEGLKKFTESLRLAYNRIG--PGVSTDDITVVT-----LSDVPSAPPQNV	629
Qy	628	VVELSNASVVDSTSMKLTW-----QIINGKYVEGPIV-----	659
Db	630	SLE-----VVNSRSIKVSWLPPPSGTQNG-FITGYKIRHKTRRGEMETLEPNLWYLF	683
Qy	660	-----YARQLPNPIVNNPAP-----VTSNT--NPL--LGSTSTSASASASALIS	701
Db	684	TGLEKGSQYSFQVSAMTVNGTGPPSNWYTAETPENDLDESQVDPQSSSLHVRPQNCIIM	743
Qy	702	T-----KPNIAAAGKRDG-----ET-----NQSG	720
Db	744	SWTPPLNPNIIVRGYIIGYGVGSPYAETVRVDSKQRYYSIERLESSHYVISLKAFNNAG	803
Qy	721	GGAP-----TPL-----	727
Db	804	EGVPLYESATTSITDPTDPVDYLLDFFTPSPDLSTPMLPPVGVQAVALTHDAVRYS	863
Qy	728	-----NTKYRMLTI--LNGGGAS-----SCITITGLVQYTYLVEFFIVPF	763
Db	864	WADNSVPKNQKTSEVRLYTVRWRTSFSASAKYKSDTSLSYTATGLKPNMYEFSVMVT	923
Qy	764	YKSEVGKPSNSRIARTLEDVPSAPYGMALLNLS--AVFLKWKAPKDRHGVLLNYH	821

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

Query Match 7.5%; Score 544.5; DB 1; Length 1040;
Best Local Similarity 21.8%; Pred. No. 5.3e-22;
Matches 253; Conservative 162; Mismatches 408; Indels 335; Gaps 49;

```

Qy      4 PRIIEHPMDTTPVK---NDPFTFNCAQEGNPTPTIQWFKDRELKTDTGSHRIMLAGGL 60
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     37 PVFEDQPLSVLPFEESTEEQVLLACRARASPPATYRWKNMNGTEMKLEPGS-RHQVLVGGNL 95

Qy     61 FFLKVIHSRRSDAGTYWCEAKNEFGVARSRNATLQVAVLRDEPRLEPANTRVAQGEVAL 120
      : : : | | | | | | | | | | | | | : | : | : | : | : | : | : | : |
Db     96 V---IMNPTKAQDAGVQQLASNPVGTTVVSREAILRFGFLQFEKSKEERDPVKAHEGWGMV 152

Qy    121 MECCAPRGSPPEQISWRKNGQTLNLVGNKRIRIV--GGNLAIQEARQSDGRYQCVVKN 178
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    153 LPCNPAPHYPGLSYRWLLN-EFPNFIPTDGRHFVSQTTGNLYIARTNASDLGNYSCLATS 211

Qy    179 VV--GTRESATAFLKVH-----RPFL---IRGPQNQTAVGVSSVVFQCRIGGDPLPDV 227
      : : : | : | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    212 HMFSTKSVFSKFAQLNLAAEDTRLFAPSIKARFPAETYALVGQQVLTLECFAGFNPPVRI 271

Qy    228 LWRRTASGGNPLRKFSWLHSASGRVHVLEDRSLKLDVDTLEDMGeytCEADNAVGGITA 287
      | | : | | : : | : : | : : | : : | : : | : : | : : | : : | : : | :
Db    272 KWRK-VDGSLSP-----QW-----TTAEPTLQIPSVSEFDEGTYECAENSKGRDITV 317

Qy    288 TGILTVHAPPKFVIRPNQLVEIGDEVLFECQANGHPRPTLYWSVEGNSLLLPgyrdGR 347
      | | | | | : : : : : : : | : | : | : | : | : | : | : | : | : | : |
Db    318 QGRIIVQAQPEWLKVISDTEADIGSLRWGCAAAKPRPTVRWLNRGE-----PLASQNR 372

Qy    348 MEVTLTPEGRSVLSIARFAREDSKVVTCNALNAVGVSSRTVVSVDTFQLPPIIEQG 407
      | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    373 VEVL-----AGDLRFESKLSLEDSG-MYOCVAENKHGTI-----YASAEIAVOALPDPRIN 422

```

```

RESULT 10
AX01_RAT
ID AX01_RAT          STANDARD;      PRT; 1040 AA.
AC P22063;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1).
GN TAX1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC TISSUE=SPINAL CORD;
RX MEDLINE=90199890; PubMed=2317872;
RA Furlay A.J., Morton S.B., Manalo D., Karagogeos D., Dodd J.,
RA Jessell T.M.;
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily

```

Query Match 7.4%; Score 540.5; DB 1; Length 1040;
Best Local Similarity 22.8%; Pred. No. 8.7e-22;
Matches 261; Conservative 160; Mismatches 442; Indels 283; Gaps 51;

```
Qy      4 PRIIEHPMDTTVPK---NDPPTFNCQAEGNPTPTIQWFKDGRKLKTDGTS-HRIMLPAGG 59
      | | | | | | | | | | | | | | | | | | | | | | | |
Db     39 PIFEEQPIGLLFPPEESAEDQVTLACRARASPPATYRWKMGNDNNLEPGSRHQQLM---GG 95
      | | | | | | | | | | | | | | | | | | | | | | | |
Qy     60 LFFLVKVIHSRRSDAGTYWCEAKNEFVGARSNRATLQVAVLRDEFRLEPANTRVAQGEVA 119
      | | | | | | | | | | | | | | | | | | | | | | | |
Db     96 --NLVIMSPYTKQDAGVYQCLASNPVGTVVSKEAVLRFGFLQEFSEKBERDPVKTHEGWGV 153
      | | | | | | | | | | | | | | | | | | | | | | | |
Qy    120 LMCEGAPRGSPPEQIISWRKNCQITLNLVGKNRIRIVD--GGNLAIQEARQSDDGRIQCVVK 177
      | | | | | | | | | | | | | | | | | | | | | | | |
Db    154 MLCPNPAPHYVGLSYRWLLN-EPNFNPTIDGRHEFVSOTGNLYIARTNASDLGNYSCLAT 212
      | | | | | | | | | | | | | | | | | | | | | | | |
```


Qy 178 NVV--GTRESATAFLKVHVR---PFLI-----RGPQNQTAVVGSSVVFQCRIGGDPLPD 226
: : | : : : : | | : | : | : : | : | :
Db 213 SHMDFSTKSVMFKFAQLNLAEDPRLFAPSIRKARFPETAYLVGQQTLECFAGFNPVPR 272

Qy 227 VLWRRITASGNNMPLRKFSLWSASGRVHLEDRSLKDDVTLEDMEYTTCEADNAVGGIT 286
: | : | : | : | : : : : | : | : | : | : | : | :
Db 273 IKWRK-VDGSLSLSP---QW-----ATAEPTLQIPSVSFEDGTYTECEAENSKGRDT 318

Qy 287 ATGILTVHAPPKVFIRPKNLQVEIGDEVLFECQANGHPRTPTLYWSEGNSSLLLPYRDG 346
: | : | : : : : : : | : | : | : | : | : | :
Db 319 VQGRRIIVQAQPEWLKVISDTEADIGSNLRWGCAAGKPRPMVRLNGE-----PLASQN 373

Qy 347 RMEVTLTPEGRSVLSIARFAREDSGKVVTTCNALNAVGSVSSRTVSVSDTQFELPPPIIEQ 406
: | : | : | : | : | : | : | : | : | : | : | : | :
Db 374 RVEVL-----AGDLRFSLKSLSDSG-MTQCVAENKHGTI---YASAEALVALAPDFRQ 423

Qy 407 GPVNQTLVPK---SIVVLCRTLTGTFVQVSWYLDGIPIDVQEHERRNLS DAGALTISDL 463
: | : | : | : | : | : | : | : | : | : | : | : :
Db 424 NPVRRILIPAARGGEISIL-CQPRAPKATILW-SKGTEI-LGNSTRVTVSDGTLIIRNI 480

Qy 464 QRHEDEGLYTCVASNRNGKSSWSGYLRDTPNPNKIFRAPELSTYPPGPKPMQVEKG 523
: | : | : | : | : | : | : | : | : | : | : | : | :
Db 481 SR-SDEGKYTCFAENFMGRANSTGILSVYRDT---KITLAPSSAD-----INVG 525

Qy 524 EN-----SVTLSTWTR---NKVGGSSLVGVVIEFMGKNETDGVAVAGTR 564
: | : | : | : | : | : | : | : | : | : | : | : :
Db 526 DNLTLCQHASHDPTMDLTFTWLDGFPIDFKPGGH-----YRRASAKETIGDLTI--- 576

Qy 565 VQNTTFTQTGLLPVGNVFFLIRAENSHGLSLSPMSEPITVGTTRYFNSGLDLSEARASLL 624
: | : | : | : | : | : | : | : | : | : | : | : :
Db 577 -----LNAHVHRG-KYTCMAQTVDGT-----SKEATVLV 606

Qy 625 SGDVVELSNASVVD--STSMKLTWQ--IINGKYVEGFYVYARQLPN---PIVNNPAPVT 676
: | : | : | : | : | : | : | : | : | : | : | : :
Db 607 RGPPGPPGPPGVVRDGTGTVQLSWSRGFDNHSPTAKYTLQARTPPSGKWKQVRTNPNIE 666

Qy 677 SN--TNPLLG-----STSTSASASASALISTK---PNIAAAGKRGETNQ 718
: | : | : | : | : | : | : | : | : | : | : | : :
Db 667 GNAETQAVLGLMPWMDYEFVRSASNILGTGEPSPGSKIRTKAEPVSVAPSGL-----S 720

Qy 719 SGGGAP-----TPLNTKYR-----MLTILNGGGAS--SCTITG-----L 750
: | : | : | : | : | : | : | : | : | : | : | : :
Db 721 GGGGAPGELIINWTPVPSREYQNGDGFGLYSPRRQSSSSQVARTVPGADAQYFVYGNDSI 780

Qy 751 VQYTLTYEFFIVPFYKSEVGKPSNSRIARTLEDVPEAPYMEALLNNSAVFLKWKAPEL 810
: | : | : | : | : | : | : | : | : | : | : | : :
Db 781 QPVTPEFVKIRSYNRRGDGPESLTALVYSAEEEPVAPAKVMAGSSSSEMNVSW-EVL 839

Qy 811 KDRHGVLLNY-----HVIRGIDTAHNFSRILTNVTIDAASPTLVLANLT 855
: | : | : | : | : | : | : | : | : | : | : | : :
Db 840 QDMGILLGYEIRYWKAGDNEAADRVRTAGLDTISAR-----VTGLN 881

Qy 856 EGVMYITVGAAGNAGVGPYCVPATLRDLPTIKRLDPFINQRDHNVDVLTQPFWFIILLGA 915
: | : | : | : | : | : | : | : | : | : | : | : :
Db 882 PNTKYHVTVRAYNRAGTG---PASPSADAMTVKPPP-----RRPPGNISWTF 925

Qy 916 ILAVLMSFGAMVFKRKHMMKQSAINTMRGNHSDVLKMPSLSARNGNGYLD----- 970
: | : | : | : | : | : | : | : | : | : | : | : :
Db 926 SSSSLSLKWDPPV-----PLRNESTVTGYKMLYQNDLHPTPLHLTSKN--WIEIPVPE 977

Qy 971 -----SSTGGMVWRPSPGGDSLEMQKDHIADYAPVCGAGSPAGGTSSSGSGGAG 1021
: | : | : | : | : | : | : | : | : | : | : | : :
Db 978 DIGHALVQIIRTG-----PGGDGIPAEV-HIV-----RNGGTSMMVESAAA 1017

Qy 1022 SGASGG 1027
: | : | : | : | : | : | : | : | : | : | : | : :
Db 1018 RPAHPG 1023

RESULT 11
DSCA_HUMAN
ID DSCA_HUMAN STANDARD; PRT; 2012 AA.

AC 060469; 060468;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE PRECURSOR (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE-BRAIN;
RX MEDLINE-98087574; PubMed-9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system.";
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20289799; PubMed-10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP FUNCTION.
RA Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
CC -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF023450; AAC17967.1; -.
DR EMBL; AF023449; AAC17966.1; -.
DR EMBL; AL163283; CAB90464.1; -.
DR EMBL; AL163282; CAB90436.1; -.
DR EMBL; AL163281; CAB90444.1; -.
DR MIM; 602523; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF000041; fn3; 6.
DR PFAM; PF000047; ig; 9.
DR PRINTS; PR00014; FNTYPEIII.
KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;

Qy 602 PITVGTRYFNSGLDLSEARASLLSGDVVELSNASVVDSTSMKLTWQIINGKYVEGFYVA 661
| | : : | | : : | | : : | |
Db 698 RIKT-----DGAAPNVAPSDV---GGGGGRNRRLTITWAPLSREYHYG----- 737
Qy 662 RQLPNPIVNNPAPVTSNTNPLLGSTSTASASASASALISTKPNIAAAGKRDTGQSGG 721
| | : : | | : : | | : : | |
Db 738 -----NN-----FGYIVAPKP-----FDGE----- 752
Qy 722 GAPTPLNTKYRMLTILNGGASCTITG-----LVQYTLYEFFVFPYKSVGEKPSN 773
: : : : | | : : | | : : | | : : | |
Db 753 -----EWKKVTVTNPD-----TGRYVHKDETMSPSTAFQVKVAFNNKGDGPYSL 797
Qy 774 SRIARTLEDVPSEAPYGEALLNNSAVFLKWKAPELKDRHGVLLNTHVIVRGIDT---- 829
: : : : | | : : | | : : | | : : | |
Db 798 VAVINSAQDAPSEAPTEVGKVLSSSEISVHWE-----HVLKIVESYQIR 843
Qy 830 ---AHNFSRLTNVTIDAASPTLVLANLTGVMYTVGVAAGNAGVGPYCPVATLRLDPI 886
| | : : | | : : | | : : | | : : | |
Db 844 YWAANDKEEAANRVQVTSQEYSARLENLLPDTQYFIEVGACNSAGCGP---PSDM-IEAF 899
Qy 887 TKRLDP-----FINQRDHF 900
| | : : | | : : | | : : | |
Db 900 TKKAPPSQPPRIISSVRSGSRYIITWDHV 928

RESULT 13

NRG_DROME

ID NRG_DROME STANDARD; PRT; 1239 AA.
AC P20241; Q24414;
DT 01-FEB-1991 (Rel. 17, Created);
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGLIAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RX MEDLINE=90030418; PubMed=2805067;
RA Bieber A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglial: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT Ll.";
RL Cell 59:447-460(1989).
RN [2]
RP SEQUENCE OF 1182-1239 FROM N.A.
RX MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bieber A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglial.";
RL Neuron 4:697-709(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RX MEDLINE=94213741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglial at 2.0 A.";
RL Neuron 12:717-731(1994).
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL
CC ADHESION IN THE DEVELOPING DROSOPHILA EMBRYO.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: NEURONS AND GLIA IN THE DEVELOPING NERVOUS
CC SYSTEM AND ON SOME OTHER NONNEURONAL TISSUES.
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M28231; AAA28728.1; ALT_SEQ.
DR EMBL; X76243; CAA53822.1; -.
DR PIR; A32579; A32579.
DR PDB; 1CFB; 30-NOV-94.
DR FLYBASE; FBgn0002968; Nrg.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 5.
DR PFAM; PF00047; ig; 6.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; 3D-structure;
KW Immunoglobulin domain; Signal; Embryo; Alternative splicing.
FT SIGNAL 1 23
FT CHAIN 24 1239 NEUROGLIAN.
FT DOMAIN 24 1138 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1139 1154 POTENTIAL.
FT DOMAIN 1155 1239 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 53 123 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 149 224 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 262 329 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 354 422 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 447 512 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 606 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 629 690 FIBRONECTIN TYPE-III.
FT DOMAIN 729 792 FIBRONECTIN TYPE-III.
FT DOMAIN 832 896 FIBRONECTIN TYPE-III.
FT DOMAIN 932 997 FIBRONECTIN TYPE-III.
FT DOMAIN 1024 1098 FIBRONECTIN TYPE-III.
FT DISULFID 59 111 POTENTIAL.
FT DISULFID 625 706
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1234 1234 T -> Y (IN REF. 2).
FT CONFLICT 1237 1237 L -> K (IN REF. 2).
SQ SEQUENCE 1239 AA; 138284 MW; 49E12692D0DD027D CRC64;

Query Match 7.3%; Score 534; DB 1; Length 1239;
Best Local Similarity 22.7%; Pred. No. 2.5e-21;
Matches 239; Conservative 135; Mismatches 438; Indels 242; Gaps 42;

Qy 4 PRIIEHP-----MDTTPKNDPFTFCQAEKNPTPTIQWFKDGRKLTDTGSHRM 54
| | : : | | : : | | : : | | : : | |
Db 29 PRITQPAPGELLFKVAQNKESDNPFIECEADGQPEPEYSWIKNGKKFDWQAYDNRL 88
Qy 55 -LPAGGLEFLKVIHSRRSDAGTYWCEAKNEFGVARSNATLQVAVLRDEFRLPANT-R 112
| | : : | | : : | | : : | | : : | |
Db 89 QPQGRGTI---VITPKDEDRGHYQCFASNEPGTATSNVTVKRAEL-NAFKDEAAKLE 144
Qy 113 VAQGEVALMECGAPRGSPPEQISW---RKNGQTLNLVGNKRIRIVDGGNLAIQEARQSD- 168
| | : : | | : : | | : : | | : : | |
Db 145 AVEGEPFLMKCAAPDGFPSPTVNMWMIQESIDGSIKNSRMTLDPEGNLWFSNVTREDA 204
Qy 169 --DGRYQCVKVN-----VGTR-----ESATFLKVHVRFPLIRGPQNTAVVGS 211
| | : : | | : : | | : : | | : : | |
Db 205 SSDFYACATSVSFRSEYKIGNKVLVDVKMGVSQNKHPVPRQYVSR--RQSLALRGK 262
Qy 212 SVVFPQCRIGGDPLDVLWRRATSGGNMPLKPKFSLWLSASGRV---HVLDRSLKLDVTL 268
: | | | : : : : | | : : | | : : | |
Db 263 RMELFCYIGGTPLPQTWVSKDG-----QRIQW---SDRITQGHY--GKSLVIRQTNF 309

```

RESULT 14
CONT_MOUSE
ID  CONT_MOUSE      STANDARD;      PRT;  1020 AA.
AC  P12960;
DT  01-JAN-1990 (Rel. 13, Created)
DT  01-JAN-1990 (Rel. 13, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  CONTACTIN PRECURSOR (NEURAL CELL SURFACE PROTEIN F3).
GN  CNTN1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6; TISSUE=BRAIN;
RX  MEDLINE=69340657; PubMed=2474555;
RA  Gennarini G., Cibelli G., Rougon G., Mattei M.-G., Goridis C.;
RT  "The mouse neuronal cell surface protein F3: a phosphatidylinositol-
RT  anchored member of the immunoglobulin superfamily related to chicken
RT  contactin."

```

```

RL J. Cell Biol. 109:775-788(1989).
CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
CC SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- MISCELLANEOUS: F3 SHARES WITH L1, N-CAM, MAG, AND OTHER CELL
CC ADHESION MOLECULES FROM NERVOUS TISSUE THE L2/HNK-1 CARBOHYDRATE
CC EPITOPE.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14943; CAA33075.1; -.
DR PIR; S05944; S05944.
DR HSSP; P40189; 1BQU.
DR MGD; MGI:105980; CNTN1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 20
FT CHAIN 21 ? CONTACTIN.
FT PROPEP ? 1020 REMOVED IN MATURE FORM.
FT DOMAIN 58 121 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 151 218 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 256 317 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 345 398 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 429 491 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 519 592 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 604 611 GLY/PRO-RICH.
FT DOMAIN 611 712 FIBRONECTIN TYPE-III.
FT DOMAIN 713 814 FIBRONECTIN TYPE-III.
FT DOMAIN 815 910 FIBRONECTIN TYPE-III.
FT DOMAIN 911 1006 FIBRONECTIN TYPE-III.
FT DISULFID 65 114 BY SIMILARITY.
FT DISULFID 158 211 BY SIMILARITY.
FT DISULFID 263 310 BY SIMILARITY.
FT DISULFID 352 391 BY SIMILARITY.
FT DISULFID 436 484 BY SIMILARITY.
FT DISULFID 526 585 BY SIMILARITY.
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1020/AA: 113388 MW: 9DCDAA40EAA4CBC7 CRC64:

```

Query Match 7.3%; Score 531.5; DB 1; Length 1020;
Best Local Similarity 22.4%; Pred. No. 2.6e-21;
Matches 234; Conservative 152; Mismatches 361; Indels 299; Gaps 45;

```

Qy      1  GENPRIIEHPMDTTVPKND---PFTFNCAQEGNPTTIIQWFKDGRELKTDGTGSHRIMLPA 57
      | | | |:| | : : ||: | :| | : : | : | :
Db      38  GFGPIFEQPIINTIYPEESLEGVSLNCRARASPPVYKWRMNGDV--DLTNDRYSMVG 95

Qy      58  GGLFFLKVIHSRRSDAGTYWCEAKNEFGVARSRNATLQAVLRLDEFRLPE-ANTRVAQG 116
      | | : : ||| |:| | |:| || || | | | : |:|
Db      96  GNLV---INNPKDKAGVYYICLASNNYGMVRSTEATLSFGYL-DPPFPEERPEVKVKEG 151

Qy      117  EVALMECGAPRSGSPES-QISWRKNQGITNLVGNKRIRIVD--GGNLAIQEAROSDDGRYO 173

```

```

RESULT 15
LAR_DROME
ID LAR_DROME      STANDARD;      PRT;  2029 AA.
AC P16621;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE DLAR PRECURSOR (EC 3.1.3.48) (PROTEIN-
DE TYROSINE-PHOSPHATE PHOSPHOHYDROLASE).

```

```

LN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046860; PubMed=2554325;
RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
RT "A family of receptor-linked protein tyrosine phosphatases in humans
RT and Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=96178473; PubMed=8598047;
RA Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S.,
RA Saito H.;
RT "The transmembrane tyrosine phosphatase DLAR controls motor axon
RT guidance in Drosophila.";
RL Cell 84:611-622(1996).
CC -!- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
CC IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
CC (PTPASE). IT CONTROLS MOTOR AXON GUIDANCE.
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
CC PIONEER NEURONS IN THE EMBRYO.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27700; AAA28668.1; -.
DR EMBL; U36857; AAC47002.1; -.
DR EMBL; U36849; AAC47002.1; JOINED.
DR EMBL; U36850; AAC47002.1; JOINED.
DR EMBL; U36851; AAC47002.1; JOINED..
DR EMBL; U36852; AAC47002.1; JOINED.
DR EMBL; U36853; AAC47002.1; JOINED.
DR EMBL; U36854; AAC47002.1; JOINED.
DR EMBL; U36855; AAC47002.1; JOINED.
DR EMBL; U36856; AAC47002.1; JOINED.
DR PIR; A36182; TDFFLK.
DR HSSP; P28827; 1RPM.
DR FLYBASE; FBgn0000464; Lar.
DR INTERPRO; IPR000242; -.
DR INTERPRO; IPR000387; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00102; Y_phosphatase; 2.
DR PFAM; PF00041; fn3; 9.
DR PFAM; PF00047; ig; 3.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00700; PRTYPHPHTASE.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW Cell adhesion; Immunoglobulin domain; Duplication.
FT SIGNAL 1 32
FT CHAIN 33 2029 PROTEIN-TYROSINE PHOSPHATASE DLAR.
FT DOMAIN 33 1377 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1378 1402 POTENTIAL.
FT DOMAIN 1403 2029 CYTOPLASMIC (POTENTIAL).

```

```

Qy 508 STYPGPPGKP---QMVEKGKNSVTLSTRNSKNVGGSSLVGYVIEMFGKNETDCGWAVGTR 564
      | | : | : : | : | | : : | | : : | | : |
Db 509 KAQQGVSPQSPNSFRATDIGETAVTLQWTKPTH--SSENIIVHY--ELYW-NDTYANQAHHR 564
      | | : | : | | | : : : : : : : | : | | | : | : |
Qy 565 VQNT-TFTQGLLPGVNYFFLIRAENSHGLSLSPMSEPI TVGTRYFNSGLDSEARSL 623
      : | : : | | | : : : : : : : | : | | | : | : |
Db 565 ISNSEAYTLDGLYPTDLYIWLAAORSQREGATTP---PIPVRTKYQVGPAPPRNITAIA 621
      | | : | : | | | : : : : : : : | : | | | : | : |
Qy 624 LSGDVVELSNASVVDSTSMKLTQWIINGKYVEGVYVARQLPNPIVNNPAPVTSNTNPLL 683
      | | : | : | : | : | : | : | : | : | : |
Db 622 TS-----STTISLSW-----LPPPV----- 636
      | | : | : | : | : | : | : | : | : | : |
Qy 684 GSTSTSASASASALISTKPNIAAAGKRDGETNQSGGGAPTPLNTKYRMLTILNGGGAS 743
      | | : | : | : | : | : | : | : | : | : |
Db 637 -----ERSNGRIIYYKVFVEVGREDDEAT-----TMTL-----NMT 668
      | | : | : | : | : | : | : | : | : | : |
Qy 744 SCTITGLVQYTYLFFIYPFYKSVBGKPSNSRIARTLEDVPSEAPYGM EALLNSSAVFL 803
      | | : | : | : | : | : | : | : | : | : |
Db 669 SIVLDELKRWYTIKIWLAGTSGVDGPRSHPIILRTQEDVPGD-PQDVKATPLNSTSIHV 727
      | | : | : | : | : | : | : | : | : | : |
Qy 804 KWKAPELKDRHGVLNLYHVIVRGI-DTAHNF-----SRLLTNVTIDAASPTLVLA 852
      | | | | : | : | : | : | : | : | : | : |
Db 728 SWKPPLEKDRNGIIRGYHIAQELRDEGKGFLNEPFKFDVVDLEFNVT----- 776
      | | : | : | : | : | : | : | : | : | : |
Qy 853 NLTEGVMTYVGVAAGNNAVG 873
      | | : | : | : | : | : | : | : | : |
Db 777 GLOPDTKYISOVAALTRKGDG 797
      | | : | : | : | : | : | : | : | : |

```

Query Match 7.2%; Score 521; DB 1; Length 2029;
Best Local Similarity 24.0%; Pred. No. 2.4e-20;
Matches 221; Conservative 131; Mismatches 359; Indels 210; Gaps 39;

```

Qy      4 PRIIEHPMDTTPVKNDPFTFNCQAEGNPTPTIQWFKDGRLEKTDTGSHRIMLPAGGLFFL 63
      | | | | : | : | | | : | : | : : : : : : | | : |
Db     36 PEIIRKPQNGVVRGGVASFYCAARGDPPPSIVWRKNGKKVSGTQSRYYTLVLEQPGGISIL 95

Qy     64 KVIHSRRSDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFLEPA-----NTRVA 114
      : : | | | | : | | | : | | : | | | | | | | |
Db     96 RIEFPVRAGRDAPYCEVAENGVGDVAVSADATLTIIY----EGDKTPAGFPVITQGFGTRI 151

Qy    115 Q-GEVALMECAGPRGSPEPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDDGRYQ 173
      : | : | | | | : | | | | | | | : | | : | | | : | : | : | : |
Db    152 EVGHTVLMTCCKA-IGNPTPNIIYWIKN-QTKVDMSNPRYSLKD-GFLQIENSREEDQSKYE 208

Qy    174 CVVKNVVGTTRESATAFLKVHVR---PFLIRGPQNTAV-VGSSVVVFCRIGGDPLPDVLW 229
      | | : | : | | | | | | | | | | : | | : | | : | | | |
Db    209 CVAENSMGTGHSKATNLVYKVRVPPPTFSRPPETISEVMLGSNNLNSCIAVSGPMPHVKW 268

Qy    230 RRTASG----GNMPLRKFSWLHSASGRVHVLEDRSLKLDVDTLEDMEGYTCEADNAVGGI 285
      : : : | | : | | : | | : | : : : | | : | : | : |
Db    269 MKGSEDLTPENEMPI-----GR-NVLQ-----LINIQESANYTCIAASTLGQI 310

Qy    286 TATGILTVHAPKFKVIRPKN-QLVEI-GDEVLFECQANGHPRPILWSVE----- 333
      : : : | : | : | : | : | : | : | : | : | : | : | : |
Db    311 DSVSVVVKYQSLP---TAPTDDQISEVTATSVRLSEYSYG-PEDLQYYIYQKPKNANQAF 366

Qy    334 GNSLLLPGRDGRMEVTLTPEGRSVLSIARFAREDSGKVVTCNA-----LNAVGSVSSR 388
      : : : | : | : | : | : | : | : | : | : | : | : |
Db    367 SEISGIITMYIVVRALSPTYETEYFTVIYVNNIGRGPSPAPACTTGETKMESAPRNQVR 426

Qy    389 TVVSVDTQFLEPPIIEQGPVNQTLPVKSIIVLPCRTLGTTPVQVSWYLDGIPIDVQEHE 448
      | : | | : | | : | : | : | : | : | : | : | : | : |
Db    427 TLSSTMYITWEPP---ETPNQGVTKYKVYY-----TTNSNOPEASW-----N 466

Qy    449 RRNLSDAGALTISDLQRHEDEGLYTCVASNRNGKSSWGYLLRLDTPTNPIKFFRAP-EL 507
      : : : | : | : | : | : | : | : | : | : | : | : |
Db    467 SQWVDNSELTYTSDVTPI---AIYT-----VRVOAYTSMGAGPMSTTPVOV 508

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:51:13 ; Search time 559.88 Seconds
(without alignments)
289.105 Million cell updates/sec

Title: US-09-540-245A-16
Perfect score: 7272
Sequence: 1 GENPRIIEHPMDTTVPKNDP.....RSLLSNSGSGTSSQPAGHNV 1381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3895.5	53.6	823	5 Q9VQ10	Q9vq10 drosophila
2	2462	33.9	859	5 Q9VP26	Q9vp26 drosophila
3	1790.5	24.6	1395	5 Q9W213	Q9w213 drosophila
4	1786.5	24.6	1395	5 O44924	O44924 drosophila
5	1498	20.6	1651	4 Q9Y6N7	Q9y6n7 homo sapien
6	1489.5	20.5	1612	11 O89026	O89026 mus musculus
7	1455.5	20.0	1651	11 O55005	O55005 rattus norv
8	1361.5	18.7	1273	5 O44928	O44928 caenorhabdi
9	1318.5	18.1	1060	11 Q9Q2I3	Q9q2i3 rattus norv
10	1256	17.3	1344	11 Q9Z2I4	Q9z2i4 mus musculus
11	1230	16.9	232	5 Q9VQ07	Q9vq07 drosophila
12	783.5	10.8	166	5 Q9VQ09	Q9vq09 drosophila
13	717	9.9	423	5 P91572	P91572 caenorhabdi
14	658	9.0	874	5 O01632	O01632 caenorhabdi
15	603	8.3	1026	11 Q62845	Q62845 rattus norv
16	602	8.3	1028	11 Q62682	Q62682 rattus norv
17	597.5	8.2	1493	11 P97798	P97798 mus musculus
18	595.5	8.2	1427	13 Q91562	Q91562 xenopus lae
19	593	8.2	1232	13 Q90284	Q90284 carassius a

20	592	8.1	1151	11 Q9QVN5	Q9qvn5 rattus sp.
21	589.5	8.1	1377	11 P97603	P97603 rattus norv
22	581	8.0	1028	11 Q07409	Q07409 mus musculus
23	579.5	8.0	1277	13 Q98902	Q98902 fuğu rubrip
24	578	7.9	1264	5 P91767	P91767 manduca sex
25	576.5	7.9	2016	5 Q9NBA1	Q9nba1 drosophila
26	575.5	7.9	1461	4 Q92859	Q92859 homo sapien
27	575.5	7.9	1461	4 O00340	O00340 homo sapien
28	575.5	7.9	2016	5 Q9V4J9	Q9v4j9 drosophila
29	572	7.9	1217	11 P97685	P97685 rattus norv
30	565.5	7.8	1272	13 Q90924	Q90924 gallus gall
31	565.5	7.8	1445	11 Q63155	Q63155 rattus norv
32	560	7.7	1443	13 Q90610	Q90610 gallus gall
33	553.5	7.6	1180	4 O15051	O15051 homo sapien
34	552.5	7.6	1280	13 Q90933	Q90933 gallus gall
35	551.5	7.6	1040	13 Q9W675	Q9w675 brachydanic
36	551.5	7.6	1302	5 O61542	O61542 drosophila
37	547.5	7.5	1166	11 Q9QVN4	Q9qvn4 rattus sp.
38	545.5	7.5	1154	11 Q9QVN3	Q9qvn3 rattus sp.
39	543.5	7.5	1018	6 Q28106	Q28106 bos taurus
40	543.5	7.5	1028	4 Q9UQ52	Q9uq52 homo sapien
41	543	7.5	1197	13 Q90478	Q90478 brachydanio
42	540.5	7.4	1028	11 P97528	P97528 rattus norv
43	538	7.4	1215	11 P97686	P97686 rattus norv
44	536.5	7.4	1822	4 Q9ULT7	Q9ult7 homo sapien
45	535.5	7.4	1028	11 Q9JMB8	Q9jmb8 mus musculus

ALIGNMENTS

RESULT 1
Q9VQ10
ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG5481 PROTEIN.
GN CG5481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beron P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kojira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003586; AAF51373.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0031341; CG5481.
DR INTERPRO; IPR001412; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 1.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PRO0014; FNTYPEIII.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN1.
SQ SEQUENCE 823 AA; 89715 MW; 36FC0B91F36F2F19 CRC64;

Query Match 53.6%; Score 3895.5; DB 5; Length 823;
Best Local Similarity 91.7%; Pred. No. 1.4e-266;
Matches 759; Conservative 0; Mismatches 2; Indels 67; Gaps 3;

Qy 11 MDTTPVKNDPFTFNCQAEGNPTPTIIQWFKDRELKTDGSHRIMLPAGGLFLKVIHSRR 70
|||||
Db 1 MDTTPVKNDPFTFNCQAEGNPTPTIIQWFKDRELKTDGSHRIMLPAGGLFLKVIHSRR 60

Qy 71 ESDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFLEPANTRVAQGEVALMECCAPRGSP 130
|||||
Db 61 ESDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFLEPANTRVAQGEVALMECCAPRGSP 120

Qy 131 EPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDGRIYQCVVNVVGTRESATAFL 190
|||||
Db 121 EPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDGRIYQCVVNVVGTRESATAFL 180

Qy 191 KVHVRPFLIRGPQNTAVVGSVVVFQCRIGGDPLVDLWRRRTASGGNPLRKFSWLHSAS 250
|||||
Db 181 KVHVRPFLIRGPQNTAVVGSVVVFQCRIGGDPLVDLWRRRTASGGNMP----- 229

Qy 251 GRVHVLEDRSLKLDVLTLEDMEYTCADNAVGGITATGILTVHAPPKFKVIRPKNLVEI 310
|||||
Db 230 GRVHVLEDRSLKLDVLTLEDMEYTCADNAVGGITATGILTVHAPPKFKVIRPKNLVEI 289

Qy 311 GDEVLFEQANGHPRTLYWSVEGNSLLLPYGRDGRMEVTLTPEGRSVLSIARFAEDS 370
|||||
Db 290 GDEVLFEQANGHPRTLYWSVEGNSLLLPYGRDGRMEVTLTPEGRSVLSIARFAEDS 349

Qy 371 GKVVTNCNALNAVGSVSSRTVVSVDTFELPPPIIEQGPVQNTLPVKSIVVLPCTRLGTPV 430
|||||
Db 350 GKVVTNCNALNAVGSVSSRTVVSVDTFELPPPIIEQGPVQNTLPVKSIVVLPCTRLGTPV 409

Qy 431 PQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDEGLYTCVASNRNGKSSWSGYLR 490
|||||
Db 410 PQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDEGLYTCVASNRNGKSSWSGYLR 469

Qy 491 LDTPTNPNIKFFRAPELSTYPPGPKPMVEKGENSVTLSTWRSNKVGGSSLVGYVIEFM 550
|||||
Db 470 LDTPTNPNIKFFRAPELSTYPPGPKPMVEKGENSVTLSTWRSNKVGGSSLVGYVIEFM 529

Qy 551 GKNETDGVAVGTRVQNTTFTQTGLLPVGNVFFLIRAENSHGLSLSPMSEPITVGT--- 607
|||||
Db 530 GKNETDGVAVGTRVQNTTFTQTGLLPVGNVFFLIRAENSHGLSLSPMSEPITVGTVSS 589

Qy 608 -----RYFNSGLDLSEARASLLSGDVVLSNASVVDSTSMKL 644
|||||

Db 590 ENHSFTLMFPLIHYFDSLPHFPQRYFNSGLDLSEARASLLSGDVVLSNASVVDSTSMKL 649
Qy 645 TW-----QIINGKYVEGFYVYARQLPNPIVNNPAP 674
|||||
Db 650 TWQVCNRLTDGSIAPHSIAHRHLIRSASFQIINGKYVEGFYVYARQLPNPIVNNPAP 709
Qy 675 VTSNTNPLLGSTSTSSASASASALISTKPNIAAAGKRDGETNQSGGGAPTPLNTKYRML 734
|||||
Db 710 VTSNTNPLLGSTSTSSASASASALISTKPNIAAAGKRDGETNQSGGGAPTPLNTKYRML 763
Qy 735 TILNGGASSCTITGLVQYTYLVEFFIVPFYKSVGEKPSNSRIARTLED 782
|||||
Db 770 TILNGGASSCTITGLVQYTYLVEFFIVPFYKSVGEKPSNSRIARTLED 817

RESULT 2

Q9VP26

ID Q9VP26 PRELIMINARY; PRT; 859 AA.

AC Q9VP26;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CG5423 PROTEIN (FRAGMENT).

GN CG5423.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; *Drosophila*.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003586; AAF51388.1; -.
DR HSSP; P56276; 1TLK.

Query Match 24.6%; Score 1790.5; DB 5; Length 1395;

Db 114 FFYRTMQGKKEQDGGYWCVAKNRVQAVSRHASLQIAVLDDFRVPEKDTVAKGETAL 173
Qy 121 MECGAPRGSPPEQISWRKNG-----QTLNLVGNKRIRIVDGNLAIQEARQSDGGRYQC 174
Db 174 LECGPPKGIPEPTLTIWIKDGVPLDDLKAMSFQASSRVIVDGNLLISNVEPIDEGNYKC 233
Qy 175 VVKNVGTRESATFLKVHVRPFLIRGPQNTAVVGSVVQCRIGGDPLDVLWRRTAS 234
Db 234 IAQNLVGTRESSYALIVQVKPYFMKEPKDQVMYLGQTATFHCSSVGGDPPKVLWKK--E 291
Qy 235 GGNMPLRKFSLHESAGRVHVEDRSLKDDVTLEDMEGYTCEADNAVGGITATGILTVH 294
Db 292 EGNIPYSRARILH-----DEKSLIENITPTDEGTYVCEAHNNVQGISARASLIVH 342
Qy 295 APPKFVIRPKNLVIGDEVLEQOANGHPRPTLYSVSEGNSSLLPGYRDRGMEVTLTP 354
Db 343 APPNFKRPSNKKVSLNGVQLPCMASGNPPSVFMTKEGVSTLFPNSSHGROYA--- 399
Qy 355 EGRSVLSIARFARESGKVVTICNALNAVGSVSSRTVVSVDQFLEPPPIIEQGPVQNTLP 414
Db 400 -ADGTLQITDVQDEGYIV-CAFSVVDSSIVRVFLQVSSVDERPPPIIQIGPANQTL 457
Qy 415 VKSIVVLCRTLGTVPQVSLVDGIPIDQEHERRNLSAGALITISLQRHEDEGLYTC 474
Db 458 KGSVAITLPCRTGNSPRIKWFHDGHA--VQAGNYSIIQSSLRVDDLQ-SDSGTYTC 514
Qy 475 VASNRNGKSSGSLRLDTPNPNKIFRAPELSTYPPGPKQPMVEKGENSVLWTR 534
Db 515 TASGERGTSWAATLVEKPGSTSL--BRAADPSTYPAPPTPKVLNVSRTISLRWAKS 572
Qy 535 NKVGGS--SLVGYVEMFGKMTDGVAVGTVRQNTTFTQGLPGVNYFFLIRAEENSHG 592
Db 573 QEKPGAVGPIIGYIYFSPDLQTGWIVAAHVRGDTQVTISGLTPTSYVFLVRAENTQG 632
Qy 593 LSLPSPMSEPTIVGTRIFN--SGLDLSEARSLSGDVVELSNASVVDSTSMKLTWQI-- 648
Db 633 ISVPSGLSNVIKTADFDAAANDLSAAR-TLLTGKSVELIDASAINASAVRLEWMLHV 691
Qy 649 -INGKYVEGFYVYRQLPNPNWNPAPVTSNTNPLLGSTTSASASASALISTKPNIA 707
Db 692 SADEKIVEGLRIHYKASVP----- 711
Qy 708 AAGKRDGETNQSGCAPTPLNTKYRMLTILNGGASSCTITGLVQTYLTYEFFIVPYKSV 767
Db 712 -----SAQYHSITVMD-ASAEFVVGNLKKYTKYEFLLTPPFETI 750
Qy 768 EGKPSNSRIARTLEDPSEAPYGEALLNSSAVFLKWKAPELKDRHGVLNLYHVIRGI 827
Db 751 EGQPSNSKTALTYEDVPSAPPDNIQIMYQTAGWVRWTPPPSQHNGNLYGYKIEV--- 807
Qy 828 DTAHNSFRILNTIDAASPTLVLANLTGVMTYVGAAGNAGVGPYCPATRLRDPIT 887
Db 808 -SAGNTMKVLNMTLNATTSVLLNLTGAVYSVRLNSFTKAGDGPYKPSISLFMDP-T 865
Qy 888 KRLDP-----FINQRDH--VNDVLTQPFILLGAI 916
Db 866 HHVHPRAHPSGTHDRHEGQDLTYHNNGNIPPGDINPTTHKKTIDYLSGFWMLVLCIV 925
Qy 917 LAVLMLSPG-AMVFKRKHMMKQ-SALNTMRGNHSDVLKMPSLSARNNGYWLDSSTG 974
Db 926 LLLVLISAISMVYFKRKBMTELKGLHLSVSDN-----EITALNINSKESLWIDHRRG 979
Qy 975 GMVWRPSPGGDSLEMQDHIADYAPVCGAPGSPAGGGTSSGSGGAGSAGSGDDINGGH 1034
Db 980 ---WR-----TADTDK-----SGLSESLLSHVN 1001
Qy 1035 GSERNQRYVGEYSNIP--TDYAEVSSFGKAPSEYGRHGNASAPYATSSILSPHQQQQ 1092
Db 1002 SSQSN-----YNSDGGTDYAEVDTNRLTFYCNCRKSPDNPTPYATMIIIGTSSSETC 1054
Qy 1093 QQPRYQRPVPGYGLQRPMPHYQQQQQQQQQAQTHQHH--QALQHQHQLPPSNIIYQM 1150

Db 1055 TKTTSISADKDS-GTHSPYSDAFAGQVPAVPVVKSNYLQYPVEPINWSEFLPP----- 1106
Qy 1151 STTSEIYPTNTGPSRSVYSEQYYPKDKQRHIH-----ITENKLSNCHTYEAAPGAKQS 1204
Db 1107 -----PPEHPPSPSTYGAQGSPESSRKSAGSGISTNQSILNASIHSSSSGGGFA 1159
Qy 1205 SPISSQFASVRRQQLPP-----NCSIGRESARFKVLNTDQGNQNLDDLGGSSMCY 1256
Db 1160 WGVSPQYAVA---CPPENVYSNPLSAVAGGTQNRQITPTNQHPQLPAY----- 1206
Qy 1257 NGLADSGCGGSPSPMAMLSHEDEHALYHTA-----DGLDDMERLYVKVDEQPPP 1307
Db 1207 --FATTGPGGAVPPNHLFATQRHAASSYQAGLNAARCAQSRACNSCDALATPSPMPPPP 1264
Qy 1308 QQQQQLIPLV-----PQHPAEGLHQLSWRNQSTRSRKNQCEIKEPSLIIYAP 1355
Db 1265 P-----VVPVPGWYQVHPNHPMHTSSNHQIY-----QCSSECS--HSR 1304
Qy 1356 GSVASERSLLSNSGSGTSSQPAHN 1380
Db 1305 SSQSHKRLQLEEHGSSAKQRGGH 1329

RESULT 5

Q9Y6N7

ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6N7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040990; AAC39575.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 20.6%; Score 1498; DB 4; Length 1651;
Best Local Similarity 28.9%; Pred. No. 9.5e-97;
Matches 410; Conservative 221; Mismatches 526; Indels 260; Gaps 42;

Qy 4 PRIIEHPMDTTPKNDPPTFNCQAEQNPPTTIQWFKDRELKTDG---SHRIMLPAGGL 60
Db 68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWKGGERVETDKDPRSHRMLPSGSL 127
Qy 61 FFLKVIHSRR-ESDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFLEPANTRVAQGEVA 119
Db 128 FFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAIRLDDFRQNPDSVMVAVGEPA 187
Qy 120 LMECGAPRGSPPEQISWRKNGQTLNLVGNKRIRI-VDGNNLAIQEARQSDGGRYQCVKN 178
Db 188 VMCEQPPRGHPPEPTISWKKDGSPLD---DKDERITIRGKLMITYRKSADGKYVCVGN 244
Qy 179 VVGTRRESATFLKVHVRPFLIRGPQNTAVVGSVVQCRIGGDPLDVLWRRTASGNN 238
Db 245 MVGERSEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRK--DDGEL 302

```

Qy      4  PRIIEHPMDTTVPKNDPFTFNCQAEGNPPTIQWFKDGRKLTDTG---SHRIMLPAGGL 60
      |||::||| | | | :| | |::||| ||||:::| | | :|| | | |::||:| | | |
Db     29  PRIVEHPSDLIVSKGEPATLNCKAERGPTPTIEWYKGGERVETDKDDPRSHRMLLPSGSL 88

Qy     61  FFLKVIHSRR-ESDAGTYWCEAKNEFVGARSNRNLTQVAVLRDEFRLEPANTRVAQGEVA 119
      |||::||| | | | | | | :| | | | | |::||:| | | | | | :| | | | |
Db     89  FFLRIVHGRKSRPDEGVYICVARNYLGEAVSHNASLEVAILLRDDPRQNPSDVMVAVGEPA 148

Qy    120  LMECGAPRGSPPEQISWRKNQGTNLNVGNKRIRI-VDGNGNLAIQEARQSDDGRYQCQVKN 178
      :||| | | | | | | | | | | :| | | :| | | | | | | | | | | | | | |
Db    149  VMECQPPRGHPPTISWKKDGSPLD---DKDERITIRGKKLMTITTRKSDAGKIVCVGTN 205

Qy    179  VVGTRESAELKXVHVRPFLIRGPQNQTAVVGSSVVFQCRIGGDPLDVLWRRRTASGGNM 238
      :|| | | | | | | | | | | :| | | | | | | | | | | | | | | | |
Db    206  MYGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRK--DDGEL 263

Qy    239  PLRKFSWLHSAAGRVHVLEDRSKLDDVTLEDMGYEYTCADNAVGGITATGILTVHAPPK 298
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    264  P-----KSRYEIRDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTVQEPFH 313

Qy    299  FVIRPKNLVEIGDEVLFECQANGHPRTILYWSVEGNSSLLL---PGYRDGRMEVILTPE 355
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    314  FVVKPRDQVVALGRVTVTQCEATGNQPAIFWRREGSQNLFSYQPPQSSRSFVSQTDG 373

Qy    356  GRSVLSIARFAREDGSKVVTNCNALNAVGSVSSRTVVSV-DTQFELPPPIIEQGPVNQTLF 414
      | | | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :|
Db    374  ---LTITNVQRSDVGYII-CQTLNVAGSIITKAYLEVIDVIADRPPIVIRQGPVNQTV 428

Qy    415  VKSIVVLPCRLTGLTPVPQVSWYLDGIPIDVQEHERRNLSDAGALTISDLQRHEDEGLYTC 474
      :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db    429  VDGTLILSCVATGSPAPTILWRKDGVLVSFTODSRIKOL-ESGVLOIR-YAKLGDTRYTC 486

```

```

RESULT 7
Q55005
ID 055005 PRELIMINARY; PRT; 1651 AA.
AC 055005;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

```

```

Qy      4 PRIEHPMDTTPVKNPDTFFNCAQEGNPTPTIQWFKDGRLEKTDTG---SHRIMLPAAGL 60
      ||::|| | | : | | |::|| |::|:| | : | | |::||:| |
Db      68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWYKGGERVETDKDDPRSHRMLPSSGSL 127

Qy      61 FFLKVIHSRR-ESDAGTYWCEAKNEFGVARSRNATLQVAVLRDEFRLPANTRVAQGEVA 119
      ||:::| | : | : | | |::|:| |::|:| |::|:| |::|:| |
Db      128 FFLRIVHGKRSRPDEGVYICVARNYLGEAVSHNASLEVALRDDFRQNPSDVMVAVGEP 187

Qy      120 LMECGAPRGSPEQPISWRKNQGTNLNVGNKRIRI-VDGGNLAIQEARQSDDGRYQCVVKN 178
      :|| | | | | | | | | | : : | : | | | | | | | | | | | |
Db      188 VMECQPPRGHPEPTISWKKGDSPLD---DKDERITIRGGKLMITYTRKSAGKYVCVGTN 244

Qy      179 VVGTRRESATFLKVHVRFLIRGPQNQTAVVGSSSVVQCRIIGDPLPDVLRWRTASGGMN 238
      :|| | | | : : | | : | | | | | | | | | | | | | | : :
Db      245 MVGERESKQVADVTLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTFGWRK--DDGEL 302

Qy      239 PLRKFSWLHSAASGRVHVEDRSKLKDDVTLEDMEGYTCEADNAVGGITATGILTVHAPPK 298
      | : : : | | : | | | | | | | | | | | | | | | |
Db      303 P-----KSRYEIRDHILKIRKVTAGDMGSYTCAENMVGKAESATLTVQEP 352

Qy      299 FVIRPKNLVLEIGDEVLFECQANGHPRTLYWSVEGNSSLLL---PGYDRGMEVTLTPE 355
      ||:::|:|:|:| | |::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      353 FVVKPRDQVVALGRVTVPQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTD 411

Qy      356 GRSVLSIARFAREDSGKVVTICNALNAVGSVSSRTVVSV-DTQFELPPPIIEQGPVQNTLP 474
      | : : | | | : | | | | : : : | | : | | | | | | |
Db      411 ---LTVTMVQRSDVGYYI-CQTLNVAGSIITKAYLEVTDVIADRPVPVIRQGPVQNTVA 467

Qy      415 VKSIVVLPCLRLGTVPVPQSWYSLWDGIPIDVQEHERRNLSAGALATISLQRHEDEGLYTC 474
      | : : | | | : | | | : | : : | : | : | | | | | |
Db      468 VDGTLTLSCVATGSPVPTILWKGKVLVSTQDSRIKQL-ESGVLIQIR-YAKLGDTRGYTC 525

Qy      475 VASNRNGKSSWSGYLRLD-----TPTNPNIKFRAPELSTYPGPPGPKQMVEKGN 525
      | | : |::| | : : | | | | : | | | | | | | | | | |
Db      526 TASTPSGEATNSAYIEVQEFQVPVQPPRPRTDPNL-----IPSASPKPEVTDVSKN 575

Qy      526 SVTLSWTRSNKVGSSSLVGYVIEFMGKNETDGGWVAVGTRVQNTTFTQTGLLPGVNYFFLI 585
      :|| | | : | : | : | | | | | | | | | | | | | | | |
Db      576 TVTLLW-QPNLNSGATPTSIIIEAFSHAGSSWQTVAVENKTTETFAIGLKPNAIYFLV 634

Qy      586 RAENSHGLSLSPMSEPI-TVGTRYFNSGDLSEARASLLSGDVV-ELNSASVVDSTSMK 643
      | | |::|:| | |::|:| | | | : : | | | | | | | | | |
Db      635 RAANAYGISDPSQISDPVKTDQVPPTTQGVHDHQVQREL---GNVVLHLHNPITLSSSSVE 692

Qy      644 LTWQI-INGKYVEGFVYAROLPNPIVNNPAPVTSNTNPLLGSTSTSASASASASALIST 702

```

Db 693 VHVTDQQSQYIQGYKILYR----- 712
Qy 703 KPNIAAGKRDGETNQSGGGAPTPLNTKYRMLTILNGGASSCTITGLVQYTYEFFIVP 762
Db 713 ----PSGASHGESEWLVFEVTRP--TK-----NSVVIDLRKGVNYEIKARP 753
Qy 763 FYKSVGEKPSNSRIARTLEDVPSEAPYGMALLL--NSSAVFLKWKAPELKDRHGVLLN 820
Db 754 FFNEFQADSEIKFAKTELEERSAPPSTVSKNDGNGTAILVTWQPPEDTQNGMVQEY 813
Qy 821 HVIVRGIDTAHNFSRILTNVTIDAASPTLVLANLTGVMYTVGVAAGNAGVGPYCVPAT 880
Db 814 KVLCLGNTRYHI----NKTVDGSTFSVIFPLVIRYSVEVAASTGAGPGVKSEPQF 868
Qy 881 LRLDPITKRLDP--FINQRDHVNDVLTQPFIIILLGAILAVLMLSGAMVFKRKHMMK 938
Db 869 IQLDSHGNPVSPEQVSLAQISDVVQPAFIAGIACWILMWSIWLRYHRK----K 924
Qy 939 QSALNTRMGNHTSDVLKMPSL-----ARNNGYWLDSSTGGMVWRPSPGGDSLEMQK 991
Db 925 RNLGSST----YAGIRKVPSTFTPTVTYQRGGEAV---SSGG---RPGLLNISEPATQ 973
Qy 992 DHIADYAPVCGAPGSPAGGGTSSGGSGGAGSG---ASGGDIHGGHSGERNQRYV--- 1044
Db 974 PWLADTWPNNGSHNDCSINCCATNGSNDSTLTYSRPADCIANYNNQDKNTLMLP 1033
Qy 1045 ----GEYSNIPTDYAEVSSFGKAPSEYGRHGNAS--PAPYATSSILSPHQQQQQQPRY 1097
Db 1034 ESTVYGDV-DLSNKNINEMKTFNSPNLKDGRFVNPSPQPTPYATQLIQANLINMNM--- 1089
Qy 1098 QQRVPVGYGLQRPMPHYQQQQHQQQAQQTQQHQAQQHQQQLPPSNYQQMSTTSEIY 1157
Db 1090 -----GGG-----DSSEKHWKPPGQQ--KQEVAPIQYNMEQNKLNKDYRANDTIL 1133
Qy 1158 PTNTGPSRSVYSEQYYPKDKQRHIIHITENKLSNCHTYEAPGAKQSSPISQFASVRRQ 1217
Db 1134 PT-----IPYN-----HSYDQMTGGSYNS--SDRGSSTSGS 1162
Qy 1218 QLPPNCSIG-RESARFKVLNTDQGNQNLDDLGGSMCYNGLADSGCGSPSPMAMLS 1276
Db 1163 Q-----GHKKGARTPKAPKQGGMNWADLL-----PPPPHPPPP 1195
Qy 1277 HEDEHALYHTADGLDD-----MERLYVKVDEQPPQQQQQLIPLVQHPAEGHLQSW 1329
Db 1196 HSNSEYSMSVDESIDQEMPCVPVPPARMYLOQDELEEEAERGPTTPVVRGAASSPAVSY 1255
Qy 1330 RNQSTRSSRKNQCEIKE-----PSELIYAP 1355
Db 1256 SHQSTATILTPSQEELQPMLODCPEDLGHPM 1286

RESULT 8

044928

ID 044928 PRELIMINARY; PRT; 1273 AA.

AC 044928;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE SAX-3.

GN SAX-3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98117250; PubMed=9458046;

RA Zallen J.A., Yi B.A., Bargmann C.I.;

RT "The conserved immunoglobulin superfamily member SAX-3/Robo directs

RT multiple aspects of axon guidance in C. elegans.";

RL Cell 92:217-227(1998).

DR EMBL; AF041053; AAC38848.1; -.

DR HSSP; P56276; 17LK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 1273AA; 139427 MW; 013E766B51A7BAD7 CRC64;

Query Match 18.7%; Score 1361.5; DB 5; Length 1273;
Best Local Similarity 28.4%; Pred. No. 2.8e-87;
Matches 357; Conservative 202; Mismatches 437; Indels 259; Gaps 37;

Qy 4 PRIIEHPMDTTPKNDPPTFNCQAEKNPTPTIQWFKDGRLE---KTDGTGSHRIMLPAGGL 60
Db 31 PVIIIEHPIDVVSRGSPATLNCAGKAPS-TAKITWYKDGQPVITNKEQVNSHRIVLDTGSL 89
Qy 61 FFLKVIHSR--RESAGTYWCEAKNEFGVARSNATLQVAVLRDEFRLPANTRVAQGEV 118
Db 90 FLLKVNKGKNGKDSAGAYYCVASNEHEGVKSNEGSLKLAILEDFFVRPRPTVQALGGEM 149
Qy 119 ALMECGAPRGSPEPQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARQSDGGRYQCVVKN 178
Db 150 AVLCSPPRGFPPEPVVSWRKDDKELRIQDMPRYTLHSDGNLIDPVDRSDSGTQCVANN 209
Qy 179 VVGTRSAFLKVVHVRPFLIRGPQNAVVGSSVVFQCRIGGDPLPDVLRRTASGGM 238
Db 210 MVGERVSNPARLSVFEPKPFQEPKDMTVDGVAFLDCRVTGDPQPIQITWKR--KNEM 267
Qy 239 PLRKFSWLHSASGRVHLED-RSLKLDVLTLEDMEYTCADNAVGGITATGILTVHAPP 297
Db 268 PVT-----RAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPP 317
Qy 298 KFVIRPKNLVIGDEVLFECQANGHPRTPLYSVEGNSLLPGY--RDGRMEVLTPE 355
Db 318 SFQKPKADQSVPAAGTATFECTLVGPSPAYFWSKEGQDLFPYSVSADGRTKVSPT-- 375
Qy 356 GRSVLSIARFAREDSGKVVTNCNALNAVGSVSSRTVSVSDTQF-----ELPPPPIEQGPV 409
Db 376 --GTLTIEVRQVDEGAYV-CAGMNSAGSSLSKAALKVTTKAVTGNTPAKPPPIEHGQ 432
Qy 410 NQTLVPKSIIVLPCRTILGTPVPQVSWYLDGIPIDVGEHRRNLSAGALITDQLRHEDE 469
Db 433 NQTLVWGSSAILPCQASGKPTPGISWLRDGLPIDTD-SRISQHSSTGLHIALKK-PDT 490
Qy 470 GLYTCVASNRNKGSSWGYLRDLTPNPNKFFRAPELSTYPGPPKQPMVEKGENSVTL 529
Db 491 GYVTCIAKNEDGESTWSASLTVEDHTS-NAQFVRMPDPSNFPSSPTQPIIVNVTDTEVL 549
Qy 530 SWTRSNKVGSSLVGYVIEFMGNETDGVAVGTRVQNTTPTQTGLLPGVNYFFLIRAEN 589
Db 550 HWNAPSTSGAGPITGYIIQYSPDLQWTFNIPDYASTEYRIKLPKSPSHSYMFVIRAEN 609
Qy 590 SHGLSLSPMSEPIVGTIFYNSGL-----DLSEARASLLSGDVVLSNASVVDSTSMK 643
Db 610 EKGIGTPSVSSALVTSKPAQVALSDKNKMDMAIAEKRLTSEQLIKLEEVTINSTAVR 659
Qy 644 LTWQIIN-GKIVEGFYVYARQLPNPVIINNPAVPTSNTPNLLGSTSTSASASASALIST 702
Db 670 LFWKKRLEELIDGYIYIKWR-----GPPRTNDNQYVNVTSPTS----- 706
Qy 703 KPNIAAGKRDGETNQSGGGAPTPLNTKYRMLTILNGGASSCTITGLVQYTYEFFIVP 762
Db 707 -----TENYVNSLMPFTNYEFFIVP 727
Qy 763 FYK---SVEGKPSNSRIARTLEDVPSEAPYGMALLNSSAVFLKWKAPELKDRHGVLLN 813
Db 728 YHSGVSHGAPSNMMDVLTAEAPPSPEDVRIIRMLNLTLRISWKAPKADGINGILKG 787
Qy 820 HVIVRGIDTAHNFSRILTNVTIDAASPTLVLANLTGVMYTVGVAAGNAGVGPY--CV 877
Db 788 FQIVIVG--QAPNNR---NITTNERAASVTLFHLVTGMTYKIRVAARSNGGVVSHGTS 842

Qy 878 PATLRDLPITKRL-----DPFIN---QRDHVNDVLTQPFIFILLGAILAVLMSFGAMV 928
Db 843 EVIMNQDILEKHLAAQENESFLYGLINKSHVP-----VIVIVAILLIFVVIIDAYC 894
Qy 929 FVK-----RKHMMKQSALNTMRGNHTSDVLKMPs-----LSARNGNGYWL 969
Db 895 YWRNSRSDGKDRSFIKINDGSVH-MASNNLWDVAQNPQNPMTAGRTMNNRNGQAL 953
Qy 970 DSST-----GGMVWRPSPGGDSLEMOKDHIADYAPVCGAPGSPAGGTSsSG 1016
Db 954 YSLTPNAQDFNNCCDYSGMTMRPG-----SEHHYHAQLTGGPGN----- 994
Qy 1017 SGGAGSGASGGDDIHGGHGSERNQQRYYGEYSNIPTDYAEVSSFGKAPSEYGRHGNASPA 1076
Db 995 -----AMSTF-----YGNQYHDDPS 1009
Qy 1077 PYATSSILSPHQ-----QQQQPRYQORPVPYGLQRPMPH-----HYQQQHQHQQA 1125
Db 1010 PYATTTLVLSNQPAWLNDKMLRAPAMPTNPVP-----PEPPARYADHTAGRRSSRA 1063
Qy 1126 QQ-----THQQHQAQQHQLPPSNI-YQQMSTISEIYPTNTGSPRSVYSEQ 1171
Db 1064 SDGRGTLNGLHHRITSGSRSDSPHTDVSYQLHSSD-----GTGSSKERTGER 1113

RESULT 9

Q9QZ13

ID Q9QZ13 PRELIMINARY; PRT; 1060 AA.
AC Q9QZ13;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance."
RL Cell 96:795-806(1999).
DR EMBL; AF182037; AAF04558.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Receptor.
FT NON_TER 1060 1060
SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C11E8542DA4 CRC64;

Query Match 18.1%; Score 1318.5; DB 11; Length 1060;
Best Local Similarity 30.8%; Pred. No. 2.3e-84;
Matches 356; Conservative 180; Mismatches 427; Indels 191; Gaps 39;

Qy 4 PRIIEHPMDTTPKNDPFTFNCQAEQNPTPTI-----QWFKDRELKTDGTGSHRM 54
Db 31 PKXVEQPEVIVSKGKPNTPNWKQKGRFPPTIGKVQRMVKGWDTKDDSKVTQG---CL 87
Qy 55 LPAGGLFLFKVIHSRR-ESDAGTYWCEAKNEGVARSNATLQAVLRLDEFLEPANTRV 113
Db 88 LPSGSLFLRLIVHGRRSKPDDEGTYVVCARNYLGEAVSRNASLEVALLRDDFRQNPDTVVV 147
Qy 114 AQGEVALMECCAPRGSPEPQISWRKNGQTLNLVGNKRIRI-VDGNNLAIQEARQSDGRY 172
Db 107 QYQRPVPGYGLQRP 1110
Db 1027 QQKSDLGFAYSLP 1040

Db 148 AAGEPAILECQPPRGHPEPTIYWKDKVRID---EKEERISIRGGKLMISNTRKSDAGMY 204
Qy 173 QCVVKNVGTRESATAFLKVHVRPFLIRGPQNTAVVSSSVFQCRIGGDLPLDVLWRRT 232
Db 205 TCVGTNMGVERDSQPAELTVFERPTFLRRPINQVLEDEPAEFRCQVQDPPQTVRWKK- 263
Qy 233 ASGGMPLRKFSLHLSASGRVHVLEDRSLKDDVLTLEDMEYTCADNAVGGITATGILT 292
Db 264 -DDADLP-----RGRYDIKDDYTLRIKKASIADEGTVCIAENRVKVEASATLT 312
Qy 293 VHAPPKFVIRPKNLVEIGDEVLFECQANGHPRTLYWSVEGNSLLPGY-RDGRMEVT 351
Db 313 VRAPPQFVVRPRDQVIAQGRVTPFCETKGNPQAVFWQKESGNLLFPNQPPQPNRSCS 372
Qy 352 LTPEGRSVLSIARFARSDSKVVTCTNALNAVGSVSSRTVSV-DTQFELPPPIIEQGPVN 410
Db 373 VSPGTG--LTITNIQRSDAGYII-CQALTVAGSILAKAQLVTVLDRPPIILQGPIN 429
Qy 411 QTLPVKSIIVLPCRTLTGTPVQVSWYLDGIPIDVQEHERRNLSAGALTISDLQRHEDEG 470
Db 430 QTLAVDGTALLKCKATG-PLPVISWLKEGFTF-LGRDPRATIQDQTLQIKNL-RISDTG 486
Qy 471 LYTCVANSRNGKSSWSGYLRDLTPTNPNKFFRAPELSTYPPGPKQVMEKGENSVTSL 530
Db 487 TYTCVATSSSGETSWSAVLDV---TESGATISKNYDTNDLPGPSKQVTDVTKNSVTSL 543
Qy 531 WTRSNKVGSSLVGYVIEFGRKNETDGMVAVGTIRVQNTTFTQTGLLPGVNYFFLIRAENS 590
Db 544 W-QGTGPGVLPASAYIIIEAFSPQSVSNWQTVANHVKTLYTVRGLRPNTIYLFMVRINP 602
Qy 591 HGLSLSPMSPEPI-TVGTRYFNSGLDLSEARASLLSGDV-VELSNASVVDSTSMKLTWQI 648
Db 603 QGLSDPSMSPDVRITQDISPPAQGVDRHQVQKEL--GDVTVRLNPNVLTPTTVQVTVTV 660
Qy 649 -INGKYVEGFVYARQLPNPIVNNPAPVTSNTNPLLGSTSTSASASASALISTKPNIA 707
Db 661 DRQPGFIQYRVYRQ-----TSGLQAST-----VWQNLDAKVPTERSAY 700
Qy 708 AAGKRDGETNQSGGAPTPLNTKYRMLTILNGGASSCTITGLVQYTYEFFIVFPYKSV 767
Db 701 LVNLKKGVT-----YEIKVRYPYNEF 721
Qy 768 EGKPSNSRIARTLEDVPEAPYMEALL--NSSAVFLKWKAPELKDRHGLLNYHVIV 824
Db 722 QGMDSKSTIRTTTEAPSAPPQSVTVLTGSHNSTSISVSWDPPADHQNGIIQYKIWC 781
Qy 825 RGIDTAHNFRIILNTVTDAASTPLVLANLTGVMYTVGAAGNNAVGPYCPATLRD 884
Db 782 LG-----NETRFHINKTVDATIRSVVIGGLEFPQIYRVEVAASTSAGVGVKSEPPQIIIG 836
Qy 885 -----PITKRLDPFINQRDHVNDVLTQPFIFILLGAILAVLMSFGAMVVKRHKM-MK 938
Db 837 GRNEVVITENNNSITEQ---ITDVVKQPAFIAGIGGACWILMGFSIWLWRRKKRGLS 893
Qy 939 QSALNTMRGNHTSDVLKMPSLSARNGNGYWLSDSTSGGMVWRPS-PG---GDSLEMQKH 993
Db 894 NYAVTFQRGD-----GGLMSNGSRPGLLNTGDP---NYPW 925
Qy 994 IADYAPVCGAPGSPAGGTSsSGGSGGAGSGAGGDDI---HGGHGSERNQQRYYGE-YSN 1049
Db 926 LADSWPATSLPVNNSNGPNEIGNFGRG-----DVLPPVPGQDKTATMLSDGAIYSS 978
Qy 1050 I-----PTDYAEVSSFGKAPSEYGRHGNASPAYATSSIL-----SPHQ-----QQQQQPR 1096
Db 979 IDFTTKTYNSSSQITQA-----TPYATTQILHSNSIHLEAVLDPDQWQSSV 1026
Qy 1097 YQRPVPGYGLQRP 1110
Db 1027 QQKSDLGFAYSLP 1040

RESULT 10
Q92214

```
ID Q9VQ07 PRELIMINARY; PRT; 232 AA.
AC Q9VQ07;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG5574 PROTEIN.
GN CG5574.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachy
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., G
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A
```

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL; AE003586; AAF51376.1; -.
DR FLYBASE; FBgn0031338; CG5574.
SQ SEQUENCE 232 AA; 25580 MW; 8EB530901DEC4EDA CRC64;

Query Match 16.9%; Score 1230; DB 5; Length 232;
Best Local Similarity 100.0%; Pred. No. 3.8e-79;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1150 MSTTSEIYPTNTGSPRSVYSEQYYYPKDKQRHIHITENKLSNCHTYEAPGAKQSSPIS 1209
Db 1 MSTTSEIYPTNTGSPRSVYSEQYYYPKDKQRHIHITENKLSNCHTYEAPGAKQSSPIS 60

Qy 1210 QFASVRRQQLPPNCSIGRESARFKVLNTDQGNQNLDDLGGSSMCYNGLADSGCGGSPS 1269
Db 61 QFASVRRQQLPPNCSIGRESARFKVLNTDQGNQNLDDLGGSSMCYNGLADSGCGGSPS 120

Qy 1270 PMAMLSHEDEHALYHTADGLDDMERLYVKVDEQPPQQQQQLIPLVPQHPAEGHLSQW 1329
Db 121 PMAMLSHEDEHALYHTADGLDDMERLYVKVDEQPPQQQQQLIPLVPQHPAEGHLSQW 180

Qy 1330 RNQSTRSSRNQGECEIKPESELIYAPGSVASERSLLNSGSGTSSQPAGHNV 1381
Db 181 RNQSTRSSRNQGECEIKPESELIYAPGSVASERSLLNSGSGTSSQPAGHNV 232

RESULT 12
Q9VQ09
ID Q9VQ09 PRELIMINARY; PRT; 166 AA.
AC Q9VQ09;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG14348 PROTEIN.
GN CG14348.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
DR EMBL; AE003586; AAF51374.1; -.
DR FLYBASE; FBgn0031340; CG14348.
DR INTERPRO; IPR001777; -.
DR PFAM; PF00041; fn3; 1.
SQ SEQUENCE 166 AA; 18353 MW; 5FDFD7163A17C217 CRC64;

Query Match 10.8%; Score 783.5; DB 5; Length 166;
Best Local Similarity 96.9%; Pred. No. 7.8e-48;
Matches 156; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

Qy 791 MEALLNNSAVFLKWKAPKELDRHGVLNHYHIVRGIDTAHNFSRLTNVTIDAASPTLV 850
Db 1 MEALLNNSAVFLKWKAPKELDRHGVLNHYHIVRGIDTAHNFSRLTNVTIDAASPTLV 60

Qy 851 LANLTEGVMTYVGAAGNAGVGPYCPATLRDLPITKRLDPFINQR-----DHVNDVLT 905
Db 61 LANLTEGVMTYVGAAGNAGVGPYCPATLRDLPITKRLDPFINQRYPINQDHVNDVLT 120

Qy 906 QPWFIILLGAILAVLMLSGFAMVFKRHHMMKQSAINTMR 946
Db 121 QPWFIILLGAILAVLMLSGFAMVFKRHHMMKQSAINTMR 161

RESULT 13
P91572
ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN ZK377.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nhan M., Hawkins J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88183; AAB52658.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00047; ig; 4.
SQ SEQUENCE 423 AA; 46544 MW; EB4530DB6BD575E5 CRC64;

Query Match 9.9%; Score 717; DB 5; Length 423;
Best Local Similarity 39.1%; Pred. No. 1.7e-42;
Matches 155; Conservative 61; Mismatches 154; Indels 26; Gaps 9;

Qy 4 PRIIEHPMDTTPVKNDFPTFNCQAEAGNPTPTIQWFKDREL---KTDGSHRIMLAGGL 60
| | | | | : | : | | | : | : | | : | : | | | |
Db 30 PVIIIEHPIDVVSRGSPATLNCAGAKPS-TAKITWYKDGQPVITNKEQVNSHRIVLDTGSL 88
| | | | | : | : | | | : | : | | : | : | | | |
Qy 61 FFLKVIHSR--RESAGTYWCEAKNEFGVARSNATLQVAVLRDEFLEPANTRVAQGEV 118
| | | : : | | | : | | | : | : | : | | | : | : |
Db 89 FLKLVNSGKNGKDSAGAYYCVASNEHEVYKSNESGLKMLREDFRVPRVTVQALGGEM 148
| : | | | | | : | | | : | : | : | | | : | : |
Qy 119 ALMECAGPRGSPPEQISWRKNGQTLNLVGNKRIRIVDGGNLAIQEARSDDGRYQCVVKN 178
| : | | | | | : | | | : | : | : | | | : | : |
Db 149 AVLCSPPRGFPPEPVVSWRKDDKELRIQDMPRYTLHSDGNLIIDPVRDSGSTYQCVANN 208
| : | | | | | : | | | : | : | : | | | : | : |
Qy 179 VVGTSATAFKLVHVRPFLIRGPQNTAVGVSSVVEFCRIGGDPLDVLWRRATSGGNM 238
| | | | | : | : | | | : | : | | | : | : |
Db 209 MVGERVSNPARLSVFEEKPFQEPKDMTVDVGAALVDFCRVTVGDQPQITWKR--KNEPM 266
| | | | | : | : | | | : | : | | | : | : |
Qy 239 PLRKFSWLHSASGRVHVLDRSLKDDVTLEDMEYTCADNAVGGITATGILTVHAPP 297
| : | : | | | : | | | | | : | : | : | : | : |
Db 267 PVT-----RAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLESAHLRVQAPP 316
| : | : | : | : | : | : | : | : | : | : | : | : |
Qy 298 KFIIRPKNLVLEIGDEVLFECQANGHPRTLYWSVEGNSLLLPY--RDGRMEVTLTPE 355

Db 317 SFQTKPAQSVPPAGGTATPCTLVGQSPAYFWSKEGQDQLFPSPVSDAGRTKVSPT-- 374
| : | : | | | | | : | | | | | | | : | : | : |
Qy 356 GRSVLSIARFAREDSGKVVTCNALNAVGSVSSRTTV 391
| : | : | | | : | : | : | : | : | : | : | : |
Db 375 --GTLTIEVRQVDEGAYV-CAGMNSAGSSLSKAAL 407

RESULT 14
O01632
ID O01632 PRELIMINARY; PRT; 874 AA.
AC O01632;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA CEESC12R.
GN ZK377.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nhan M., Hawkins J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88183; AAB52657.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 1.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 874 AA; 95861 MW; BC72270818D734C9 CRC64;

Query Match 9.0%; Score 658; DB 5; Length 874;
Best Local Similarity 22.9%; Pred. No. 7.6e-38;
Matches 229; Conservative 154; Mismatches 329; Indels 288; Gaps 35;

Qy 400 PPPIIEQGPVNTLPVKSIIVLPCRTLTGPVQVSWYLDGIPIDVQEHERRNLSAGALT 459
| | | | | : | : | | | : | : | | | : | : | : | : |
Db 28 PPPIIEHGHQNTLMVGSSAILPCQAGKPTPGISWLRDGLPIDITD-SRISQHSHTGSLH 66
| : | : | : | : | : | : | : | : | : | : | : | : |
Qy 460 ISDLQRHEDEGLYTCVASNRNGKSSWSGYLRDLTPTNPIKFFRAPELSTYPPGPKRPM 519
| : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 15
Q62845
ID Q62845 PRELIMINARY; PRT; 1026 AA.
AC Q62845;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE NEURAL CELL ADHESION PROTEIN BIG-2 PRECURSOR.
GN BIG-2.
OS Rattus norvegicus (Rat).

```

```

Qy      2 ENPRIEHPMDTTPVKNDPFTFNCQAEGNPPTTIQWFKDGRKLKTDGT-SHRIMLPAGGL 50
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     30 QEPESHVMFPLDSEEEK---VKLSECVKGKPKPHIRWKLNGTDV--DIGMDFRISVVEGSL 84
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy     61 FFLKVIHSRRSDAGTYWCEAKNEFGVARSRNATLQVAVLREDFRLEPANT-RVAQGEVA 119
      : : : : | : | | : | | | | | | | | : | : | : | : | : | : |
Db     85 L---INNPKTQDSGTQCIATNSFGTIVSREAKLQFAYL-ENFKTRTRTSVSVRRQGQM 140
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    120 LMECGAPRGSPEPQISWRKNQGTNLNVGNKRIRIVDGGNLAIQEARQSDDGRIYQVVKNV 179
      : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    141 VLLCGPPHSGELSYAWIFN-EHPSYQDNRRFVSQETGNLYIAKVEKADVGNYTCVVTNT 199
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    180 VGTRESATAFLKVHVRPFLIRG-----PQNQTAVVGSSVVFQCRIGGDPL 224
      : : : : : : : | : | : | : | : | : | : | : | : | : | : |
Db    200 VTSHQ-----VLGPPTPLILRNDGVMGEYEPKIEVQFPETVPAEKGSTVKLECFALGNPV 254
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    225 PDVLWRRTASGGMPLRKFSWLHSASGRVHVLEDRSLKDDVTLIEDMGEYTCADNAVGG 284
      : | : | | | | : | : : : | : | | : : : | : | : | : | : | : |
Db    255 PTILWRR-ADG--KPIARKARRHKSSG-----ILEIPNQQEDAGSYECVAENSRGK 303
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    285 ITATGILTVHAPKPEVIRKPNQLVEIGDEVLFEQANGHPRTPLYWSVEGNSLLLPGYR 344
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    304 NIARKQVTFYAQPNWQIINDIHVAMESVFWECKANGRPKPTYRWLNKGD-----PLLT 358
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    345 DGRMEVTLTPEGRSVLSIARFAREDSGKVVTCNALNAVGSVSSRTVSVSDTFQLPPPII 404
      : : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db    359 RERIQIE-----QGTLNITIVNLSDAG-MYQCVAENKHGVIIASAELSV-----IA 403
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    405 EQGPVNQTL-----PVKSIIVLPCRTLTPVPQVSWYLDGIPIDVQEHERRNLSDAGA 457
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    404 ESPDFSRTLLKRVTLVKVGGEVIECKPKASRPVYTW-RKGREI-LRENERITISEDGN 461
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    458 LTISDLQRHEDEGLTYCVASNRNCKSSWSGYRLDPTPTNPNI-----499
      | | : : : | | | | : | : | : | : | : | : | : | : | : | : |
Db    462 LRIINVTK-SDAGSYTCIATNHFGTASSTGNVVKDPTKVMVPPSSMDVTVGESIVLPCQ 520
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    500 -----499
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    521 VTHDHSLDIVETWTFNHGLIDFDKGDHFERVVGQDSAGDLMIRNIQLKHAGKYVCMVQT 580
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy    500 ---KFRAPELSTYPGPGKQPMV---EKGENSVTLSWTRSNKVGSSSLVGVIEM---F 550
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    581 SVDKLAAADL-IVRGPPGPPRAVITDEIITDTTAOLSW-RPGDNDHSPITMVIQARTPF 638
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Search completed: January 22, 2001, 12:52:00
Job time: 1921 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:17:58 ; Search time 233.01 Seconds
(without alignments)
190.332 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYLGFYHTHTHTYINFD.....TAQRFRSIPRNGIYEQET 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*

- 1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6860	100.0	1297	20	Y13565	C. elegans Robo po
2	6860	100.0	1297	20	Y08403	C. elegans ROBO pr
3	1588	23.1	1395	20	Y13563	Drosophila Robo 1
4	1588	23.1	1395	20	Y08401	Drosophila sp. ROB
5	1500.5	21.9	1651	20	Y13566	Human Robo 1 polyp
6	1489.5	21.7	1649	20	Y08404	Human ROBO1 protei
7	1350	19.7	1380	20	Y08402	Drosophila sp. ROB
8	1344.5	19.6	1381	20	Y13564	Drosophila Robo 2
9	1266	18.5	753	20	W83927	Human T85 protein.
10	600.5	8.8	1018	18	W06485	Rat contactin liga
11	600	8.7	1571	19	W42087	Human Down syndrom
12	598	8.7	1910	19	W42086	Human Down syndrom

13	593	8.5	1018	15	R63759	Human contactin (F
14	593	8.5	1018	17	R87028	Human contactin.
15	583.5	8.5	1257	20	W74152	Human L1 cell adhe
16	583	8.5	1028	19	W29667	Homo sapiens DL195
17	580.5	8.5	1225	19	W52289	Homo sapiens cdo t
18	561.5	8.2	434	20	Y13567	Human Robo 2 polyp
19	561.5	8.2	434	20	Y08405	Human partial ROBO
20	554	8.1	1447	16	R68553	Deleted in colorec
21	554	8.1	1447	20	Y33498	Human DCC protein.
22	554	8.1	1728	12	R13144	Deleted in Colorec
23	546.5	8.0	1242	19	W52287	Rattus norvegicus
24	541.5	7.9	1192	19	W57900	Protein of clone C
25	539.5	7.9	1299	21	Y40439	Human Nr-CAM prote
26	526	7.7	1897	21	Y81785	Human protein tyro
27	526	7.7	1897	21	Y56100	LAR tyrosine phosph
28	524.5	7.6	1304	19	W59994	Human neural cell
29	518.5	7.6	1251	19	W37778	Rattus norvegicus
30	510.5	7.4	1496	20	W81030	Melanoma associate
31	510.5	7.4	1496	21	Y70469	Human p53 target m
32	508.5	7.4	4412	21	Y53666	Sequence gi/101742
33	501.5	7.3	3117	21	Y53667	Sequence gi/332818
34	495	7.2	1911	16	R71726	Human PTP-OB. Hom
35	495	7.2	1911	18	W27225	Human protein tyro
36	495	7.2	1911	20	W94027	Human protein tyro
37	485.5	7.1	1139	19	W37779	Rattus norvegicus
38	485	7.1	1125	19	W52288	Rattus norvegicus
39	465	6.8	1501	16	R72858	Rat receptor type-
40	441.5	6.4	1070	18	W08747	Human colon carcin
41	439	6.4	761	17	R92255	Neural cell adhesi
42	434	6.3	1853	21	Y53668	Protein 608 sequen
43	434	6.3	2387	21	Y53665	Mechanical stress
44	434	6.3	2597	21	Y53664	Mechanical stress
45	432	6.3	848	21	Y88565	Human NCAM 140kd i

ALIGNMENTS

RESULT 1
Y13565
ID Y13565 standard; Protein; 1297 AA.
XX
AC Y13565;
XX
DT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 1297 AA;

Query Match 100.0%; Score 6860; DB 20; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYLGFYHTHTHTTYINFDKIPNASNLAPVIIIEHPIDVVSRGSPATLNCGAKPSTAKI 60
DB 1 mylgyfhththttyinfdkipnasnlapviiiehpdivvsvrgspatlnccgkpkstaki 60
61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKVNKGKNGKSDAGAYVCVASNEHGEVKS 120
DB 61 twykdqgpvitnkeqvnshrivldtgsifllkvnsgkngkdsdagayvcvasnehgevks 120
121 NEGSLKLAMLRDFRVRPRTVQALGEMAVLECSPPRGFPPEVSVWRKDKELRIQDMPR 180
DB 121 negslklamrdrfrvrprtvalgcmavlecspprgfppevsvwrkddkelriqdmpr 180
181 YTLHSDGNLIIDPVDSDSGTYQCVANNMVGERSVNPRLSVFEKPKFEQEPKDMTVDVG 240
DB 181 ytlhsdgnliidpvdrsdsgtyqcvannmvgersvnparslvfekpkfeqepkdmtdvvg 240
241 AAVLDFCRVTGDPQGITWKRKNEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300
DB 241 aavldfcrvtgdpqgitwkrknempvtrayiakdnrglriervqpsdegeyvcyarnp 300
301 AGTLEASAHLYVQAPPSFQTKPADQSVPAAGTATFECTLVGQPSPAYFWSKEGQDQLLF 360
DB 301 agtleasahlyvqappsftkpadqsvpaggtatfectlvqgpspayfwskegqddllf 360
361 SYVSADGRKVSPTGTLTIEEVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKK 420
DB 361 syvsadgrtkvsptgtltieevrqvdegayvcagmnsagslskaalkatfetkgrvqkk 420
421 KSKMGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGHNQTLMVGSAILPCQASGKP 480
DB 421 kskmgkqkqknvqsiikylisavtgntpakppptiehgqntlmvgssailpcqasgkp 480
481 TPGISWLRDGLPIDITDSRISQHSSTGSLHIALKPKDTGVYTCIAKNEDGESTWSASLT 540
DB 481 tpgiswlrldglpiditdsrisqhsstgslhialkpkdtgvtyciaknedgestwsaslt 540
541 EDHTSNAQFVRMPDPSNFPSSPTQPIIVNVTDTEVELHWNAPSTSGAGPITGYIIQYSP 600
DB 541 edhtsnaqfvrmpdpsnfpssptqpiivnvtdtevelhwnapstsgagpitgyiiqysp 600
601 DLGQTFWNPIDYVASTERYIKGLKPSHSYMFVIRAENEKIGTSPVSSALVTTSKPAQV 660
DB 601 dlgtqfwnpidyvasteryikglkpshsymfviraenekigtspvssalvttskpaqv 660
661 ALSDKNMDMAIAEKRLTSEQLIKLEEVKTINSTAVRLFVKKRLEELIDGYIKWRGPP 720
DB 661 alsdknmdmaiaekrltseqlikleevktinstavrlfwwkrleelidgyikwrgpp 720
721 RTNDNQYVNVTSPTSTENYVVSMLPFTNYEFFVPIYHSGVSHIHGAPNSMDVLTAEAPP 780
DB 721 rtndnqyvntspststenyvvsmlpftnyeffvpiyhsghshihgapsmdvltaeapp 780
781 SLPPEDVRIRMLNLTTLRISWKAPKADGKILGKQFQIVIGQAPNNRNITTNERAASV 840
DB 781 slppedvrirmlnltrlriswkapkadgkilkqfqiwigqapnnrnittneraasv 840

QY 841 TLFHLVTGMTYKIRVAARSNGGVGSHGTSEVIMNQDTLEKHLAAQGENESFLYGLINKS 900
DB 841 tlfhlvtgmytkirvaarsnggvshgtsevimnqdtlekhlaaqgenesflyglinks 900
901 HVPVIVIVAILIIFVVIILAYCYWRNSRSDGKDRSFIKINDGSHMASNNLWDVAQNPN 960
DB 901 hvpvivivailiifvviilaycywrnsrsgdkdrsfikindgshmasnnlwdvaqnpn 960
961 QNPMYNTAGRMTMNNRNGQALYSITPNAQDFFNNCDDYSGTMHRPGSEHHYHAQLTGGP 1020
DB 961 qnpyntagrmtmnnrngqalysitpnaqdfnnccddysgtmhrpgsehyhaqltggp 1020
1021 GNAMSTFYGNQYHDDPSPYATTTLVLNQPAWLNDKMLRAPAMPTNPVPEPPARYADH 1080
DB 1021 gnamstfygnqyhddpspyatttlvlnsqpawlndkmlrapamptnpvpepparyadh 1080
1081 TAGRRSRSSRASDGRGTLNGGLHRTSGSQRSDSPHTDVSQVQLHSSDGTGSSKERTGE 1140
DB 1081 tagrrsrssrasdgrgtlngglhrtsgsqrdspphtdvsyqlhssdgtgsskertge 1140
1141 RRTPPNKTLMDFIPPPSPNPPPGGHVYDTATRRQLNRGSTPREDTYDSVSDGAFARVDV 1200
DB 1141 rrtppnktlmdfipppspnpppgghvydtatrrqlnrgstpredtydsvsdgafarvdv 1200
1201 NARPTSRNRNLGGRPLKGRKDDSDQRSSLMDDGGSSEADGENSEGDPVPRGGVRKAVPR 1260
DB 1201 narptsrnrnlggrplkgkrddsdqsrsslmddggsseadgensegdpvprggvrkavpr 1260
1261 MGISASTLAHSCYGTNGTAQRFRSIPRNNGIVTQEQT 1297
DB 1261 mgisastlahscygtngtaqfrsiprnnngivtqeqt 1297

RESULT 2

Y08403

ID Y08403 standard; Protein; 1297 AA.

XX

AC Y08403;

XX

DT 24-JUL-1999 (first entry)

XX

DE C. elegans ROBO protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX

OS Caenorhabditis elegans.

XX

PN W09920764-A1.

XX

PD 29-APR-1999.

XX

PF 20-OCT-1998; 98WO-US22164.

XX

PR 14-NOV-1997; 97US-0971172.

PR 20-OCT-1997; 97US-0062921.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman CS, Kidd T, Mitchell KJ, Tear G;

XX

DR WPI; 1999-312615/26.

DR N-PSDB; X57252.

XX

PT Robo polypeptides, a new immunoglobulin superfamily member

XX

PS Claim 1; Page 59-63; 80pp; English.

XX

CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful

CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.

XX
SQ Sequence 1297 AA;

Query Match 100.0%; Score 6860; DB 20; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYILGFYHTHTHTYINFDKIPNASNLAPVIEHPIDVVVSRGSPATLNCGAKPSTAKI 60
Db 1 myilgfyhththtyinfdkipnasnlapviehpidvvvsgspatlncgakpstaki 60

Qy 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKVNKGKSDAGAYYCVASNEHGEVKS 120
Db 61 twykdqgpvitnkeqvnsshrivldtgsflfkvnsgksgdsgayyvcasnehgevs 120

Qy 121 NEGSLKLAHLREDFVRPRTVQALGGEMAVLECSPPRGFPPEVSWRKDDKELRIQDMPR 180
Db 121 negslklamhredfvrprtvaqalggemavlecspprgfppevswrkddkelriqdmpr 180

Qy 181 YTLHSDGNLIIDPVRSDSGTYQCVANMVGERSVNPRLSVFEKPKFEQPKDMTVDVG 240
Db 181 ytlhsdgnliidpvrdsdsgtyqcvanmvgerstnparlsvfeqpkfepkdmtdvvg 240

Qy 241 AAVLFCRVTGDPQPIITWKRKNEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300
Db 241 aavlfcrvtgdpqpqitwkrknempvtrayiakdnrglriervqpsdegeyvcyarnp 300

Qy 301 AGTLEASHLRVQAPPSFQTKPADQSVPAAGTATFECTLVGQSPAYFWSKEGQDQLLF 360
Db 301 agtleasahlrvqappsfgtkpadqsvpaggtatfectlvqgspayfwskegqddllf 360

Qy 361 SYVSADGRTKVSPGTGLTIEEVQRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQK 420
Db 361 syvsadgrtkvsptgtltieevrqvdegayvcagmnsagslskaalkatfetkgrvqk 420

Qy 421 KSKMGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGQNTLMVGSSAILPCQASGKP 480
Db 421 kskmgkqkqknvqsiikylisavtgntpakppptiehgqntlmvgssailpcqasgkp 480

Qy 481 TPGISWLDGLPIDITDSRISQHSSTGSLHIALKKPDTGVYTCIAKNEDGESTWSASLT 540
Db 481 tpgiswldglpiditdsrisqhsstgslhialkkpdtgvtyciaknedgestwsaslt 540

Qy 541 EDHTSNAQFVRMPDPSNFPSSPTQPIIVNVTDEVELHWNAPSTGAGPITGYIIQYSP 600
Db 541 edhtsnaqfvrmpdpnsfpssptqpiivnvtdevelhwnapstgagpityiiqysp 600

Qy 601 DLGQTFWNPIDYVASTERYIKGLKPSHSYMFVIRAENEKGIGTPSVSSALVTTSKPAQV 660
Db 601 dlgtwfnpidyvasteryikglkpsysmfviraenekgigtspvssalvttskpaqv 660

Qy 661 ALSKKNKMDMAIAEKRLTSEQLIKLEEVTINSTAVRLFVKKRLEELIDGYIKWRGPP 720
Db 661 alsdknkdmaiaekrltseqlikleevktinstavrlfkwkrleelidgyikwrgpp 720

Qy 721 RTNDNQYVNTSPSTENYVVSNLMPFTNYEFFVIPPYHSGVHSIHGAPNSMDVLTAEAPP 780
Db 721 rtndnqyvntspstenyvvsnlmpftnyeffvipyhsgvshihgapsnmdvltaeapp 780

Qy 781 SLPPEDVRIRMLNLTIRISWKAPKADGINKGFIQIVIGQAPNNRNITTNERAASV 840
Db 781 slppedvrirmnltilriswkapkadgingikgfivigqapnnrnttneraasv 840

Qy 841 TLFHLVTGMYKIRVAARSNGGVGSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKS 900
Db 841 tlfhlvtgmykirvaarsnggvshgtsevimnqdtlekhlaaqenesflyglinks 900

Qy 901 HVPVIVIVAILIIFVVIIAICYWRNSRNSDGKDRSFIKINDGSVHMASNNLWDVAQNP 960
Db 901 hvpvivivailiifvviiaicywrnsrnsdgkdrsfikindgsvhmasnnlwdvaqnp 960

Qy 961 QNPMYNTAGRTMNNRNGQALYSITPNAQDFFNCCDDYSGMTMRPGSEHHYHYAQLTGGP 1020
Db 961 qnpyntagrtmnnrngqalysitpnaqdfnccddysgmtmrpgsehyhyaqtlgtgp 1020

Qy 1021 GNAMSTFYGNQYHDDPSYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADH 1080
Db 1021 gnamstfygnqyhdpsyatttlvlsnqpawlndkmlrapamptnpvpepparyadh 1080

Qy 1081 TAGRRSRSSRASDGRGTLNGGLHRTSGSQRSDSPPHDVSIVQLHSSDGTGSSKERTGE 1140
Db 1081 tagrrssrsrasdgrgtlngglhrtsgsqrdspphdtvsyqlhssdgtgsskertge 1140

Qy 1141 RRTPPNKTLMDFIPPPSPNPPPGHVYDTATRLNLRGSPREDTYDSVSDGAFARVDV 1200
Db 1141 rrtppnktlmdfipppspnpppgghvydtatrlnlrgstpredtydsvsdgafarvdv 1200

Qy 1201 NARPTSRNRLGGRPLKGRDDDSQRSSLMDDGGSSEADGENSEGDPVPRGVRKAVPR 1260
Db 1201 narptsrnrlggrplkgkrrdddsqrsslmddggsseadgensegdvprrgvrkavpr 1260

Qy 1261 MGISASTLAHSCYGTNGTAQFRSIPRNGIVTQEQT 1297
Db 1261 mgisastlahscygtngtaqfrsiprningivtqeqt 1297

RESULT 3

Y13563

ID Y13563 standard; Protein; 1395 AA.

XX

AC Y13563;

XX

DT 30-JUL-1999 (first entry)

XX

DE Drosophila Robo 1 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
modulation; nerve cell function.

XX

OS Drosophila sp.

XX

PN W09925833-A1.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24327.

XX

PR 14-NOV-1997; 97US-0065543.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55767.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX

PS Disclosure; Page 30-33; 56pp; English.

XX

CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX

SQ Sequence 1395 AA;

Query Match 23.1%; Score 1588; DB 20; Length 1395;
Best Local Similarity 31.0%; Pred. No. 2.1e-89;

Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

Qy 29 APVIEHPIDVVSRGSPATLNC--GAKPSTAKITWKDQGPVITNKEQVNSHRIVLDTG 86
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 55 sprliehptdlvvkknepatlnckvegkpept-iewfkdgpevstnekk--shrvqfkdg 111

Qy 87 SLFLLKVNKGKNGKSDAGAYYCVASNEHEGVKSNESGLKLAHLREDFRVRPRVQALGG 146
: | | : | | : | | | | | : | | : | | | | | : | | : | | : | | : | |
Db 112 alffyrtnmgk--keqdggeywcavknrvqgavsrhaslqlavlrdffrvepkdtrvakg 169

Qy 147 EMVLECSPPRGFPPEPVVSWRKDD-----KELRIQDMPRYTLHSDGNLIIDPVDSDSG 200
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 170 etallecpgpkipeptliwkdqgvplddlkamsfgassrvrivdgnllinsvpeiddeg 229

Qy 201 TYQCVANVMGERSVSNPARLSVFEPKPEQEKDPTVDVGAHLFDCRVTDGDPQITWK 260
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 230 nykclaqnlvgstressyakkilvqkpyimkpekdqvmlyggqtatfhcsvsgdppkvlwk 289

261 RKNEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPPSQT 320
: : | | | | : | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 290 keegnipsrarihlhdeksleisnitptdegtyvceahnvvgqisaraslihvappnftk 349

Qy 321 KPADQSVPAAGGTATFECTLVGQSPAYFWSKEGQDLFPSTVSADGRTKVSPGTGLTIE 380
: | | | : | | : | | : | | | | | : | | : | | : | | : | | : | | : | |
Db 350 .rpsnkkvlgngvqlpcmasgnppsvfwtkegvstlmfpn--sshgrqyvaadgtlqit 407

Qy 381 EVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKQKNVQSIKLYI 440
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 408 dvrqedegeyyvcsafsvvdsstvrflqvs----- 437

Qy 441 SAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRDGLPIDITDSRI 500
: | | : | | : | | | | | : | | : | | : | | : | | : | | : | |
Db 438 -----svderpppiiqigpangtlpkgsvatlpcratgnpsrikwfhghavq-agnry 491

Qy 501 SQHSTGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFP 560
: | | : | | : | | | | | : | | : | | : | | : | | : | | : | |
Db 492 siiqgsslrddlqlsdsgtytctasgergetswaatltvekgpgts-lhraadpstypa 550

Qy 561 SPTQPIIVNVTDEVELHW--NAPSTSGAGPITGIIQYSPDLGQTFWNPIDYVASTEY 618
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 551 ppgtpkvlvnrtsislrwaksqekpgavgiigtyveyfspdltgwivaahrvgdtqv 610

Qy 619 RIKGLKPSHSYMFVIRAENEKIGTSPVSSALVTTSKPAAQVALSDKNKMDMAIAEKRLT 678
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 611 tisqltpgtsyvlvraentqgisvpsglsvniktieadfaasan----dlsaartllt 666

679 SEQLIKLEEVTINSTAVRLFVKRKL--EELIDGYIKWRGPPRTNDQY--VNVTPS 734
: : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 667 gks-velidasainasavrlewmhvsadekyveglrihyk-dasvpsaqyhsitvmdas 724

Qy 735 TENYVVSNLMPPTIEFFVPIYHSGVHSINGAPSNSMDVLTAEPSPDPEDVRIRMLNL 794
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 725 aesfvgnlkkkykyeffltpf--fetiiegpsnsktaltiedvpsappndiqigmyq 781

Qy 795 TTLRISWKAPKADGINGILKGFQI-VIVGQAPNNRNITTNERAASVTLFHLVTGMTYKI 853
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 782 tagwvrtpppsqhngnlygykiesagntmkvianmtlnattsvllnlttgavysv 841

Qy 854 RVAARSGNGGVGSHGTSEVIMNQDTLEKHLA-----AQQENESFLYGLINKSHVP- 903
: : : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 842 rlmsftkagdgypskpislfdm-ptphvhpprahpsgthdgrhegqdltyh--nngnipp 898

Qy 904 -----VIVIVAILIIFVVIYIAYCYWRNSRNSDGRSFIKIND 942
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 899 gdinptthkkttdylsgpwlmlvclivilvlvisaaismvyfkrkhqmtkelghlsvvsd 958

Qy 943 GSVHMASNN---LWDVAQNPNQNPMTYTAGRMTMNNRNQALYSLTPNAQDFNNCDDY 998
: : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 959 neitalninskeslw----idhhrgwrtadtdkdsqleskllshvnssqsnynnsd-- 1011

Qy 999 SGTMRPGSEHHYHAQLTGPGNAMSTFYG-NQYHDDPSPIATTLV----- 1045
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1012 ggt-----dyaev---dtrnlttfynckrpsndptpyattmigtsssetctkt 1058

Qy 1046 -----LSNQPPA--WLNDKMLRAPAMPTN-----PVPPE--PPARYAD 1079
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1059 sisadksgthspysdafagqvpavpvvksnylqypvepinwseflpppphepppsstyg 1118

Qy 1080 HTAG--RRSRSSRASDGRG-----TLNGLLHRTSGSQRS-----DSPHTDVS 1122
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1119 yaqgspeessrksksagsglistngsilnasihsssggfsawgvsqyavacppenvysn 1178

Qy 1123 VQLHSSDGTGSSKERTGERRTPPNKTLMDFIPPPSPNPPPGGHVYDTATRRQ----- 1175
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1179 plsavaggtqnryqitptnqhppqlpay-fattgpggavpp-nhl-pfatqrhaaseyqa 1235

Qy 1176 -LNRGSTPREDTYDS-----VSDGAFARVDVNA---RPTSRNRNL--- 1211
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1236 glnaarcacsracnsdcalatpsmpqppppvpypegvyqpvhpnshpmhtssnhqiyqc 1295

Qy 1212 GGRPLKGRDDDSQRSSLMDDGGSSEADGENSEGDVP 1250
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1296 ssecsdhsrsgshkrqlgleehgssakqrgghrrrap 1334

RESULT 5

Y13566

ID Y13566 standard; Protein; 1651 AA.

XX

AC Y13566;

XX

DT 30-JUL-1999 (first entry)

XX

DE Human Robo 1 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout; modulation; nerve cell function.

XX

OS Homo sapiens.

XX

PN WO9925833-A1.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24327.

XX

PR 14-NOV-1997; 97US-0065543.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55770.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX

PS Disclosure; Page 44-48; 56pp; English.

XX

CC The invention relates to a method for modulating the amount of Comm (commissureless) polypeptide in contact with a cell expressing active Robo (roundabout) on its surface. The method comprises modulating the effective amount of Comm polypeptide in contact with the cell, where the amount of expressed active Robo is specifically modulated inversely with the modulation of the effective amount of Comm in contact with the cell. The method is used to modulate the amount of active Robo expressed on a cell. The method can be used to screen for agents that modulate Robo:Comm interactions. This is particularly useful for modulating nerve cell function.

XX

SQ Sequence 1651 AA;

Query Match 21.9%; Score 1500.5; DB 20; Length 1651;

Best Local Similarity 32.1%; Pred. No. 6.5e-84;

Matches 408; Conservative 166; Mismatches 480; Indels 219; Gaps 36;

Qy 267 PVT-----RAYIAKDNRGLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPP 316

Db 239 pirkfswlhasgrhvwled-rsklddvtledmgeytceadnavggitatgiltvhapp 297
Qy 317 SFQTKPADQSVPAAGGTATFECTLVGPSPAYFWSKEGQDLPSPYSVADGRTKVSP-- 374
Db 298 kfvirpknqlveigdevlfecqanghrptlywsvegnsslllpgy--rdgrmevtltp 355
Qy 375 --GLTIEEVRQVDEGAYV-CAGMNSAGSSLSK--AALKATFETKGRVQKKSKMGKQKQ 429
Db 356 grsvlsiarfaredsqkvvtcnalnnavgsvsrtvsvdtqfel----- 399
Qy 430 KNVQSIKYLISAVGNTPAKPPPTIEHGQNTLMVGSSAILPCQASGKPTPGISWLRD 489
Db 400 -----pppieggpvnqtlpvksivlpcrtlgtvpvqvsyld 438
Qy 490 GLPIDITD-SRISQHSSTGLHIADLKK-PDTGVYTCIAKNEDGESTWSASLTVEDHTS-N 546
Db 439 gipidvqeherrnlsdaqaltisdlqrhedeglytcvasnrrngksswgylrldtptn 498
Qy 547 AQFVRMPDPSNEPSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLQQTW 606
Db 499 ikffrapelstypgpgkpmvekgensvltswtrsnkvsgslvgviemfgknetdgw 558
Qy 607 FNIPDYVASTERYIKGLKPSHSYMFVIRAENKEGIGTPSVSSALVTTSKPAAQVALSDKN 666
Db 559 vavgtrvntfttqglpgvnyffliraenshglslpspmsepitvgrtynsfl---- 614
Qy 667 KMDMAIAEKRLTSEQLIKLEEVKTIINSTAVRLFVKKRLEELIDGYIKWR----- 717
Db 615 --dlsearassllsgdvvelsnasvvdtsmklwtqiln-gkyvegfyryarqlpnivnn 671
Qy 718 -GPPRTNDNQYVNVTSPTS----- 734
Db 672 papvtstntpllgststsasasasalistkpniaaagkrdgetngsggaptplntky 731
Qy 735 -----TENYVVSNLMPFTNYEFFVPIYHSGVHSIHGAPNSMDVLTAEAPPSLPP 784
Db 732 rmltilngggassctitglvqytlveffivpfyk---svegkpsnsriartledvpseap 788
Qy 785 EDVRIRMLNLTTLRISWKAPKADGINGILKGFQIVIG--QAPNNR---NITTNERAAS 839
Db 789 ygeallnssavflkwkapelkdrhgyllnyhvlrvrgidtahnfirllntvtdaaspt 848
Qy 840 VTLFHLVTGMTYKIRVAARSNGGVSGHGTSEVIMNQDTLEKHLAAQENESFLYGLINK 899
Db 849 lylanltgvmtyvgvaagnnagvgpy--cvpatrlrldpikr-----dpfin---qr 897
Qy 900 SHVP-----VIVIVAILIIFVIIAYCYWRNSRSDGKDRSFIKINDGSVH-MASN 950
Db 898 dhvndvltqpwfiillgailavlmlsfamvfvk-----rkhhmmkqsalntrng 948
Qy 951 NLWDVAQNPNQNPMTAGRTMNNRNGQALYSLTNPQAQDFNCCDYSGMTMRPG---- 1006
Db 949 htsdvlkmps-----lsarngngyldsst-----ggmvrwpspggd 985
Qy 1007 -----SEHHYHYAQLTGGPCN----- 1022
Db 986 slemqkdhiadyapvcgagsgaggtstsgsggagsgagsggddihgggsernqrryvg 1045
Qy 1023 -----AMSTF-----YGNQYHDDPSYATTLVLSNQPAWLNDKMLRAPAMPT 1066
Db 1046 eysniptdyaevsfgkapseygrhgnaspapayatsllsphqq-----qqqqqpyqyq 1099
Qy 1067 NPVP-----PEPPARYADHTAGRRSRSSRASDGRGTLNGLHRTSGSQRSDSPHTDV 1120
Db 1100 rpvpgyqlgrpmhp-----hyqqqqhqqqqagq-----thqhqalqhqqlpsni 1146
Qy 1121 SYVQLHSSD-----GTGSSKERTGER-RTPNKTLMDFIPPPSPNPPPGHGYVDYATR 1174
Db 1147 -yqqmsttseiyptntgprsvsyseqyyppdkqrhihienklsn-----chtyeaapga 1200
Qy 1175 QLNRGSTPREDTYDSVSDGAFARVDVNARPTSRNRNLGRPLKKG--RDDDSQRSSLMMD 1232

Db 1201 ---kqsspissqfas-----vrrqqlppncsigresarfvlntdggknqnnld 1247
Qy 1233 DDGGSSEADGENSEGDVPRGGVRKAVPRMGISASTLA-----HSCYGT 1275
Db 1248 ldgssmcyngladsg-----cggspspnamimshedehtlyht 1285

RESULT 8

Y13564

ID Y13564 standard; Protein; 1381 AA.

XX

AC Y13564;

XX

DT 30-JUL-1999 (first entry)

XX

DE Drosophila Robo 2 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.

XX

OS Drosophila sp.

XX

PN W09925833-A1.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-US24327.

XX

PR 14-NOV-1997; 97US-0065543.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;

XX

DR WPI; 1999-338008/28.

DR N-PSDB; X55768.

XX

PT Modulation of Robo-Comm polypeptide interactions

XX

PS Disclosure; Page 34-38; 56pp; English.

XX

CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX

SQ Sequence 1381 AA;

Query Match 19.6%; Score 1344.5; DB 20; Length 1381;

Best Local Similarity 26.8%; Pred. No. 2e-74;

Matches 386; Conservative 216; Mismatches 486; Indels 353; Gaps 46;

Qy 30 PVIEHPIDVVSRRGSPATLNCGAKPS-TAKITWYKDGQPVITNKEQVNSHRIVLDTGSL 88
Db 4 priehpmdtvpkndpftncqaegnptptiqwkdqrel---ktdtghrmlpaggl 60
Qy 89 FLLKVNNGKNGKSDAGAYTCVASNEHGEVKSNEGLKMLREDFRVRPRTVQALGGEM 148
Db 61 fflkvihsr--resdagtywceaknefvgsrnatlqvavlrdefrlepantrvaggev 118
Qy 149 AVLECSPPRGFPVSVWRKDDKELRIQDMRYTLHSDGNLIIDPDRSDSGTYQCVANN 208
Db 119 almeagaprgpepgiswrknqgtlnlvgnkrlivdgynlaigeargsddgryqcavkn 178
Qy 209 MVERVSNPARLSVFEKPKFEQPKDMTVDVGAALFDCRVTGDPQITWKR--KNEPM 266

```
Qy 1165 GHVYDTATRRQLNRGSTPREDTYDSVSDGAFARVDVNARPTSRNRNLGGRPLKGG--RDD 1222
      | | : | : | : | : | : | : | : | : |
Db 1192 chtyeaapga---kqsspissqfas-----vrrqglppncsigresarfkvintd 1238

Qy 1223 DSQRSSLLMDDGGSSSEADGENSEGDVPRGGVRKAVPRMGISASTLA-----HSCYG 1274
      : : | | : | : | : | : | : | : | : |
Db 1239 gqknqgnllldlgssmcyngladsg-----cggspspmamlmshedehalyh 1285

Qy 1275 T 1275
      |
Db 1286 t 1286
```

XX
BT
Hess isolated human HTLV-III and -IV containing isolates in 1983.

PT New isolated human FTHMA-070 and T85 proteins - used to develop

Query Match 8.8%; Score 600.5; DB 18; Length 1018;
Best Local Similarity 23.7%; Pred. No. 7.5e-29;
Matches 255; Conservative 124; Mismatches 374; Indels 323; Gaps 39;

Qy 4 LGFYHTH-THTHTYINFDKIPNASNLAPVIIIEHPIDVVVSRS---PATLNCAGKAPSTAK 59
Db 19 lfeftwhrryghvseedk----gfgpfieeqpintiypeeslegkvslnccrarsapfp 73
Qy 60 IYWK---DGQPVITNKEQVNSHRIVLDTGSLFLKVNKSGKNGKSDAGAYYCVASNEH 115
Db 74 v--ykrwmngdvdltm-----drysmvggnlvi-----nnpdkqkdagiyyclasny 120
Qy 116 GEVKSNEGSLKAML---REDFRVRPRTVQALGEMAVLECSPPRGFPPEPV-VSWRKDD 170
Db 121 gmvrsateatlsfgyldpfpped---rpe-vkvkegkqmvllcdpphyfddlsyrwnef 176
Qy 171 KELRIQDMPRYTLHSDGNLIIDPVRSDSGTYQCVANNMVGERSVNP-RLSVFEK--- 225
Db 177 pvfitmdkrrfvsqtnqnylianvessdrngyscf-----vsspsitksvskfipl 228
Qy 226 ---PKFEQEP-----KDMTVDVGAALVFCRVTGDPQPIQITWKRKNEMPVTRAYI 273
Db 229 ipiperttkypadivvqfkdiymmgqntlecfalgnpvpdirwkvlempmtt-aei 287
Qy 274 AKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPPSPQTKPADQSVPAAGGTA 333
Db 288 stsgavlkifnqledeglyceaeenirgdkhqariyvqafpewvehindtevdigsd 347
Qy 334 TFECLTVGPSPAYFWSKEGQDQLLFPYSVSDGRTKVSPTGTLTIEEVQVDEGAYCA 393
Db 348 ywpcvatgkpiptirwlkngy-----ayhkgelrlydvtfenagmyqci 391
Qy 394 GMNSAGSSLSKALK---ATFETKGRVQKKSKMGKQKQKVNQSIIRYLISAVTGNT 448
Db 392 aenaygtiyanaelkilalaptfe-----mnpmkkk-----ilaa----- 426
Qy 449 AKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWRLDGLPIDITDSRISQHSGL 508
Db 427 -----kgrrvileckpkapkpksfswk-gtewlvnssrillwedgsl 468
Qy 509 HIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHT-----SNAQFVRM 552
Db 469 einnitrdgglytcfenngkanstgtlvtinptriilapinaditvgenatmcqaas 528
Qy 553 PDPS-----NF----- 558
Db 529 fdpsldltfwsfngyvidfnkeithiygrnfmldangellirnaqlkhagrytctaqt 588
Qy 559 -----PSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDL 602
Db 589 ivdnssasadlvrrgppgppglriedratsvaltsrgsdnhs-piskytiq-kttil 646
Qy 603 QOTWFI---PDYVASTERYIKG---LKPSHSMYFIRAENEKIGTSPVSSALVTISKPA 657
Db 647 sddwkdaaktdpiiegmesakavdlipmeyerfvvatntltgtgepsipsnriktgaa 706
Qy 658 AQVALSD-----KNKMDMAIAEKRLTSEQLIKLE----- 686
Db 707 pnvapsdvvggggtareltitwapsreyhyngnfyivafkpfdgeewkkyvtvntpdg 766
Qy 687 -----EVKTINSTAVR 697
Db 767 ryvhketntpstafqkvkafnnkgdgyysliavinsaqdapseaptevqkvlsseis 826
Qy 698 LFWKKRLEELIDGYIYKWRGPPRTNDNQYVNVSTSPSTENTV---VSNLMPTTNYEFFVIP 755
Db 827 vhw-khvlekiyesyqiriyaghdkeaaahrvqvt---qeysarlenltpdtqyfieva 882
Qy 756 YHSGVSHIGAPNSMDVLTAEAPSLPPEDVRIRMLNLTTLR-----ISWKAPKADGI 809
Db 883 cns---agcgpssdvietftrkapsqpp---ri---issvrsrgsyiitwdhvalsn 932
Qy 810 NGILKGFQIVIGQAPNNRNITTNERAASVTLFHLVTGMTYKIRVAARSNGGVGV 865
Db 933 estvtgkilyrpdqghdglfsthksievp---iprdgeyvvevrahsdggdgv 985

RESULT 11

W42087

ID W42087 standard; Protein; 1571 AA.

XX

AC W42087;

XX

DT 28-SEP-1998 (first entry)

XX

DE Human Down syndrome-cell adhesion molecule DS-CAM2.

XX

KW DS-CAM2; Down syndrome-cell adhesion molecule; neural cell;

KW signal transduction; trisomy 21; mental retardation;

KW holoprosencephaly; corpus callosum agenesis;

KW schizencephaly; diagnosis; assay; human.

XX

OS Homo sapiens.

XX

PN WO9817795-A1.

XX

PD 30-APR-1998.

XX

PF 23-OCT-1997; 97WO-US19547.

XX

PR 25-OCT-1996; 96US-0029322.

XX

PA (CEDA-) CEDARS SINAI MEDICAL CENT.

XX

PI Korenberg JR;

XX

DR WPI; 1998-271791/24.

DR

N-PSDB; V31988.

XX

PT New isolated Down's Syndrome-cell adhesion molecule - used to

PT develop products for detection, diagnosis and therapy of

PT developmental and neurological abnormalities

XX

PS Claim 2; Page 90-95; 109pp; English.

XX

CC This polypeptide comprises Down syndrome-cell adhesion molecule

CC DS-CAM2, an extracellular soluble protein belonging to a novel

CC subclass of the Ig superfamily with highest homology to neural cell

CC adhesion molecules. Its amino acid sequence was deduced from cDNA

CC clones (see V31982) isolated from a trisomy 21 foetal brain library.

CC It is a splice variant of membrane-bound DS-CAM1 (see W42086), and

CC lacks the entire transmembrane domain of DS-CAM1. The invention

CC provides human and murine DS-CAM nucleic acid sequences (see also

CC V31981, V31985-87), expression vectors and host cells, transgenic

CC animals, antibodies, antisense oligonucleotides, and primers

CC derived from DS-CAM nucleic acids. DS-CAM polypeptides are associated

CC with developmental and neurological processes. They can be used in

CC e.g. neural prosthetic devices used in entubulation methods of

CC repairing (regenerating) damaged or severed peripheral nerves, and

CC also in bioassays to identify agonists and antagonists. The products

CC can also be used in detection, diagnosis and therapy of developmental

CC and neurological abnormalities such as Down syndrome, mental

CC retardation, holoprosencephaly, agenesis of the corpus callosum,

CC or schizencephaly.

XX

SQ Sequence 1571 AA;

Query Match 8.7%; Score 600; DB 19; Length 1571;
Best Local Similarity 25.1%; Pred. No. 1.4e-28;
Matches 228; Conservative 129; Mismatches 373; Indels 180; Gaps 38;

Qy 16 YINFDKIPNASNLAPVIIIEHPIDVVVSRS---PATLNCAGKAPSTAKITWYKDG 66

Db 386 fvrkdcl-saqdyvqvldgtptkilsafsekvvpaepvslmncvkgtpititwldd 444

Qy 67 QPVITNKEQVNSHRIVLDTGSLFLKVNKSGKNGKSDAGAYYCVASNEHGEVKSNEGSLK 126

Db 445 dpilkggshrisqmitsegnvlsyniss---sqvrdggvyrctannasagv----- 493

```

RESULT 12
W42086
ID W42086 standard; Protein; 1910 AA.
XX
AC W42086;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human Down syndrome-cell adhesion molecule DS-CAM1.
XX
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis; assay; human.
XX

```

OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..23
FT		/label= Sig_peptide
FT	Protein	24..1910
FT		/label= Mat_protein
FT	Domain	24..887
FT		/label= IG
FT		/note= "immunoglobulin type-C2 domain"
FT	Domain	888..1594
FT		/label= FbN
FT		/note= "fibronectin type III domain"
FT	Domain	1595..1616
FT		/label= Transmembrane
FT	Domain	1617..1910
FT		/label= Cytoplasmic
FT	Region	24..126
FT		/label= Ig1
FT	Region	127..225
FT		/label= Ig2
FT	Region	226..316
FT		/label= Ig3
FT	Region	317..409
FT		/label= Ig4
FT	Region	410..506
FT		/label= Ig5
FT	Region	507..603
FT		/label= Ig6
FT	Region	604..697
FT		/label= Ig7
FT	Region	698..792
FT		/label= Ig8
FT	Region	793..887
FT		/label= Ig9
FT	Disulfide-bond	46..102
FT	Disulfide-bond	145..197
FT	Disulfide-bond	246..293
FT	Disulfide-bond	335..385
FT	Disulfide-bond	428..484
FT	Disulfide-bond	525..575
FT	Disulfide-bond	617..669
FT	Disulfide-bond	711..766
FT	Disulfide-bond	809..865
FT	Disulfide-bond	1307..1359
FT	Modified-site	78..80
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	106..108
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	470..472
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	487..489
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	658..660
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	666..668
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	710..712
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	748..750
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	795..797
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	924..926
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	1142..1144
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	1160..1162
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	1250..1252
FT		/note= "Asn is N-glycosylated"
FT	Modified-site	1271..1273
FT		/note= "Asn is N-glycosylated"

```

Qy      16 YINFDKIPNASNLAPVIIIEHP-----DVVVSRSGPATLNCAGKPS-TAKITWYKDG 66
      :: ||: :| : ||:| : ||| :|: || : ||| |
Db      386 fvrkdcl-saqdyvgvledgtpklisafsekvsvpaeypslmcnvkgtpitptitwtdld 444

Qy      67 QPVITNKEQVNSHRI---VLDTGSFL-LLKVNSGKNKGSDAGAYCVASNEHEGVKSNE 122
      ||: |||| : ||: | :| : | | | | :| | |
Db      445 dpilkg---gshrisqmitsegnvsvynliss---sqvrdggvyrctannasagv--- 493

Qy      123 GSLKLA MLREDFR---VRP-RTVQALGGEMAVLECSPPRGFPPEPVSWRKDDKELRIQD 177
      | | : | :|| : : | : | :| : | | :| : | |
Db      494 ----lyqarinvgpasirpmknitalagrdrtylhc-rvigypysikwyknsnlipfh 548

Qy      178 MPRYTLHSDGNLIIDPVDRS-DSGTQVCVANMVGVERSVN-----PARLSVFEPK 226
      : :| :| : | : | | | | | :| :| : | | |
Db      549 r-qvafenngtliklsdvqkevdegeytc--nvlvqpqlstsgsvhvtkvvpfiqpfefp 605

Qy      227 KFEQEPKDMTVDVGA AVLFC- RVTGDPQPQITWKRKNEPMVTRAYIAKDN---RGLR 281
      : :| :| : | : | | | | | :| :| : | | |

```

```

RESULT 13
R63759
ID R63759 standard; Protein; 1018 AA.
XX
AC R63759;
XX
DT 10-MAY-1995 (first entry)
XX
DE Human contactin (EMBL Accession #221488).
XX
KW Human contactin; human brain glycoprotein; neural cell adhesion;
KW mouse F3, chicken contactin/F11 adhesion molecules.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 45..94
FT Disulfide-bond ; 138..191
FT Disulfide-bond 243..290
FT Disulfide-bond 332..371
FT Disulfide-bond 416..464
FT Disulfide-bond , 506..563
FT Domain , 604..657
FT /label= FLR
FT /note= "Conserved core of fibronectin type III
FT ; like repeat"

```

CC is homologous to mouse F3 and chicken contactin/F11 adhesion
CC molecules; all three are involved in neural cell adhesion. An
CC antibody to HBGC can be used in an immunoassay to detect HBGC in
CC samples, and the cDNA can be used as a probe to detect a sequence
CC encoding HBGC in samples.

XX
SQ Sequence 1018 AA;

Query Match 8.6%; Score 593; DB 15; Length 1018;
Best Local Similarity 23.5%; Pred. No. 2.2e-28;
Matches 246; Conservative 126; Mismatches 361; Indels 316; Gaps 36;

```
Qy      30 PVIIIEHPIDVVVSRGS---PATLNCGAKPSTAKI-TWKDQGQPVITNKEQVNSHRIVLDT 85  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db     41 pifeeqpintiypeeslegkvslnccrarpfpvykwrmnngdv----dltsdrysmvg 95  
  
Qy     86 GSLFLKLVKNCGNKGDSDAGAYICVASNEHEGVKSNEGSKLKAML-----REDFVRVP 138  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db    96 gnlvi-----nnpdkqkdagilyyclasnnygmvrsteatlsfygldpfppeerpevrkve 150  
  
Qy   139 RTVQALGGEMAVLECSPPRGFPPEPV-VSWRKDDKELRI-QDMPRYTLHSDGNLIIDPVDR 196  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   151 -----gkgmvllicdpphyfipddlsrywllnefpvitmdkrkfvsqtgnlyianvea 203  
  
Qy   197 SDSGTICYCVANNMVGERSVNP-RLSVFEK-----PKFEQEP-----KDMTVDV 239  
       |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   204 sdkgnyscf-----vsspsitksvfskfiplpiperttkypypadivvgfkdvayalm 255  
  
Qy   240 GAAVLFDCRVTDGPDPQITWKRKNEMPVTRAIYIAKDNRGLRIERVQPSDEGEYVCYARN 299  
       | |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   256 gqnvtlecfalgnvpvdirwrklempst-aelstsgavlkifniqldegelyceaeen 314  
  
Qy   300 PAGTLEASAHLRVQAPPSFOTKPADQSPVAGGTATFECTLVGQSPAYFWFSKEGQQDLLF 359  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   315 irgdkhqhar:yqvafpewehndtevdigsdlywpcvatgkpiptirwlkngy----- 369  
  
Qy   360 PSYSADGRYKVSPTGLTIEEVROVDEGAYVCAGMNSAGSSLSKAALK-----ATFETK 414  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   370 -----ayhkgeirlvydtfenagmyqciaentygaiyanaekilalaptfe-- 416  
  
Qy   415 GRVQKKSKMKQKQKNVQSIIKYILISAVTGNTPAKPPPTIEHGHNQTLMVGSSAILPC 474  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   417 -----mnpmkkk-----lla-----kggrvliec 436  
  
Qy   475 QASGKPTPGISWLRLGPLIDTDSRSIQHSTGLSHIADLKKPDTGVYTICAKNEDGESTW 534  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   437 kpkaapkpksfswk-gtewlvnssriliwedgsleinnitrdnggiytcfaaennrgkans 495  
  
Qy   535 SASLTVEDHT-----SNAQFVRMPDPS----- 556  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   496 tgtlvtidptriilapinaditvgenatmqcaasfdpaldltfwsvngyvidfnkenih 555  
  
Qy   557 ---NF-----PSSPTQPIIVNV 570  
       ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   556 yqrnfmdnsqellirnlaqlkhagrytctaqtivdnssasadlvrrpgppgggriedl 615  
  
Qy   571 TDTEVELHWNAPSTSGAGFITGYIIQ---YYSPDLQQTWFNIPDYASTE-YRIKLKPS 626  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   616 ratsvaltwrgsdnhs-piskytiqtktildwdwkadtppiegnmeearavdlipw 674  
  
Qy   627 HSYMFIARAENEKGIGTPSVSSALVTTSKPAQVALSD----- 664  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   675 meyerfvratntlgrgepsipsnrktdgaapnvapsdvvgggggrnreltitwapsrey 734  
  
Qy   665 --NNKMMAIAEKRLTSEQLIKLE----- 686  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   735 hygnngfyiafakpfidgewkvvtvnptdgryvhkdetmspstaftgvkvkfannkgdgp 794  
  
Qy   687 -----EVKTINSTAVRLFWRKKRLEELIDGYI-K-WRGPPRTND 724  
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db   795 vsllavinsacdapseapteevqkylsseisvhw-ehvlekivesvaivrwaahdekea 853
```

Query Match 8.6%; Score 593; DB 17; Length 1018;

[illegible]

Qy 288 SDEGEYVCYARNPAGTLEASAHLRVQAPPSFQTKPADQSVQAGGTATFECTLVGQPSPAY 347
|:|:| | | | | : : : | : | : | | : | : | : |
Db 305 eddgyrciaenslgssarhayvtveaapywlhkpqshlypggetarldcqvggrpqpev 364
Qy 348 FWSKEGQDQLLPSPYSADGRKVSPTGTLTIEEVRQVDEGAYVCAGMNSAGSSLSKAAL 407
| | : : : | : : | | : | : | : | : | : | : |
Db 365 twring---ipveelakdqyriq-rgalilsnvqpsdtmvtqcearnrhglllanayi 419
Qy 408 KATFETKGRVQKKKSKMGKQKQNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMV- 466
| : : | | | | | | | | | |
Db 420 -----yvv-----qlpakilta-----dnqymav 439
Qy 467 -GSSAILPCQASCKPTPGISWL-RDGLPIDITDSRISQHSGLHIADLKKPDTGVYTCI 524
|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 440 qgstayllckafgapvpsqwldeggtv-lqderffpyangtlgirdlqandtgryfcl 498
Qy 525 AKNEDGESTWSASLTVEDHTSNAQFVRMP-----DPSNFP----- 560
| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 499 aandqnnvtimanlkvkdatqitqgprstiekkgsvrtftcqsfdpslqpsitwrgdgr 558
Qy 561 -----SPTQPIIVN----- 569
| | : |
Db 559 dlqelgdsdkyfriedgrlvihslqysdqqnyscvasteldvvesraqlllvsgpgpvprl 618
Qy 570 -----VTDTEVELHWNAPSTSGAGPITGYIIQYISPDLG-QTWFN---IPDYVASTEY 618
:|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 619 vlsdlhlitqsgvrsvw-spaedhnapekydiefedkemapekwyslgkvpqngtsttl 677
Qy 619 RIKGLKPSHSMFVIRAENEKIGITPSVSSALVTTSKPAQVALSDKNKMD----- 669
: | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 678 k---lspvyhytfrvtainkygpgpspsvsetvvtpe----aaepknvdpvkgegnett 729
Qy 670 -MAIAEKRLT-----SEQLI----- 683
| | | | | : : |
Db 730 nmvitwklrwmwnapqvyrvqwrpggtrgpwqegivsdplvsvntstfvpvyeikvq 789
Qy 684 -----KLEEVKTINSTAVRLFVKRRKLEEL---IDGY 713
:|:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 790 avnsqgkgepgqvtigysgedypqalpelegieilnssavlvkwrpvdlaqvkgghlrgyn 849
Qy 714 IK-WR-GPPRTNDNQYVN----VTSPSTENYVVSNLMPFTNYEFFVPIPYHSGVHSIHGAP 767
: | | | | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 850 vtywregsqrkshrhkhkhvvpanttsvilsgrlryss-----yhlevqafngrg 902
Qy 768 SNSMDVLTAEAPSLP--PEDVRIRMLNLTTLRISWKAPKADGINGILKGFQIVIGQAP 825
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 903 sgaseftfstpegvpghealhlecqsnstlllrwqpplsh--ngvltgyvlsyhlde 960
Qy 826 NNNRNITTNERAASV---TLFHLVTGMTYKIRVAARSNGGVG 864
: : | | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 961 ggkgqlsfnlrdpelrthnldlsphlryrfqlqattkegpg 1002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:25:18 ; Search time 325.28 Seconds
(without alignments)
270.743 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYLGFYHTHTHTHTYINFD.....TAQRFRSIPRNNIGVTQEQT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6523.5	95.1	1273	2	T42405	sax-3 protein - Ca
2	4628	67.5	874	2	T29548	hypothetical prote
3	2232	32.5	423	2	T29549	hypothetical prote
4	1505.5	21.9	1612	2	T30805	duttl protein - mo
5	1483.5	21.6	1651	2	T14160	transmembrane rece
6	1323	19.3	1344	2	T14316	rig-1 protein - mo
7	635	9.3	1028	2	I58164	BIG-1 protein - ra
8	632.5	9.2	1443	2	I50600	neogenin - chicken
9	616	9.0	1028	2	A53449	plasmacytoma-assoc
10	607	8.8	1375	2	T13822	frazzled gene prot
11	598	8.7	1896	2	T08851	Down syndrome cell
12	597	8.7	1010	2	JU0094	Fil protein precur
13	595	8.7	1020	2	S05944	neuronal cell surf
14	595	8.7	1021	2	A57112	contactin precurs
15	593.5	8.7	1091	2	S01998	contactin precurs
16	593	8.6	1018	2	A54744	contactin 1 precu
17	592	8.6	1018	2	JC4211	neural adhesion pr
18	589	8.6	1272	2	S26180	neurofascin - chic
19	588	8.6	2222	2	T13924	sdh protein - frui
20	585	8.5	1040	2	A34695	axonal glycoprotei
21	583.5	8.5	1257	1	A41060	neural cell adhesi
22	582.5	8.5	1040	2	A49356	transient axonal g
23	580.5	8.5	1240	2	T03097	CDO protein - huma
24	580.5	8.5	1427	2	I51669	tumor suppressor -
25	574	8.4	1259	2	A43425	Bravo/Nr-CAM cell
26	567.5	8.3	1260	1	S05479	neural cell adhesi
27	562.5	8.2	1036	2	S22383	axonin 1 precursor
28	561	8.2	1268	1	A39640	neural cell adhesi
29	557	8.1	1259	2	S36126	neural cell adhesi

30	554	8.1	1447	2	A54100	tumor suppressor p
31	553	8.1	1277	2	T30532	neural cell adhesi
32	552.5	8.1	1526	2	T13823	frazzled gene prot
33	550	8.0	1239	1	A32579	neuroglian - fruit
34	547	8.0	1256	2	T03096	CDO protein - rat
35	531.5	7.7	1197	2	T30581	neural cell adhesi
36	527	7.7	1907	2	S50893	protein-tyrosine-p
37	526	7.7	1897	1	TDHULK	leukocyte antigen-
38	518.5	7.6	1898	2	S46216	leukocyte antigen-
39	509.5	7.4	1912	2	A56178	protein-tyrosine-p
40	508.5	7.4	7962	2	I38346	elastic titin - hu
41	501.5	7.3	5175	2	T20992	hypothetical prote
42	501.5	7.3	5198	2	T43290	hemictin precurs
43	501	7.3	1894	2	C54689	protein-tyrosine-p
44	498.5	7.3	1232	2	T43027	neural cell adhesi
45	492	7.2	1863	2	S46217	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T42405
sax-3 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42405
R:Zallen, J.A.; Yi, B.A.; Bargmann, C.I.
Cell 92, 217-227, 1998
A>Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple
A:Reference number: 222160; MUID:98117250
A:Accession: T42405
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <ZAL>
A:Cross-references: EMBL:AF041053; NID:g2804779; PIDN:AAC38848.1; PID:g2804780
C:Genetics:
A>Note: sax-3
C:Function:
A>Description: sax-3 function is required at the time of axon guidance

Query Match 95.1%; Score 6523.5; DB 2; Length 1273;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1243; Conservative 1; Mismatches 1; Indels 33; Gaps 3;

Qy	24	NASNLAPVIEHPIDVVS	RGSPATLNC	GAKPSTAKITWYKDGQPVITNKEQVNSHRIVL	83
Db	25	DASNLAPVIEHPIDVVS	RGSPATLNC	GAKPSTAKITWYKDGQPVITNKEQVNSHRIVL	84
Qy	84	DTGSLFLKVN	SGKNGKSDAGAIYCVASNEHGEVKSNEGSLKLA	MLREDFRVRPRIVQA	143
Db	85	DTGSLFLKVN	SGKNGKSDAGAIYCVASNEHGEVKSNEGSLKLA	MLREDFRVRPRIVQA	144
Qy	144	LGEMAVLECSPPRGFP	PEPVVSWRKDDKELRIQDMPRYTLHSDGNLIIDP	VDRSDSGTYQ	203
Db	145	LGEMAVLECSPPRGFP	PEPVVSWRKDDKELRIQDMPRYTLHSDGNLIIDP	VDRSDSGTYQ	204
Qy	204	CVANNMVGERSNP	PARLSVFEKPKFQEPKDMTVDVGAAVLFDCRV	TGDPQPIITWKRKN	263
Db	205	CVANNMVGERSNP	PARLSVFEKPKFQEPKDMTVDVGAAVLFDCRV	TGDPQPIITWKRKN	264
Qy	264	EMPVTRAYIAKDN	RGLRIERVQPSDEGEYVCYARNPAGTLEASAH	LRVQAPPSFQTKPA	323
Db	265	EMPVTRAYIAKDN	RGLRIERVQPSDEGEYVCYARNPAGTLEASAH	LRVQAPPSFQTKPA	324
Qy	324	DQSVPAAGTATFECT	LVGQPSPAYFWSKEGQQLLFPSPVSADGR	KVSPGTGLTIEEVR	383
Db	325	DQSVPAAGTATFECT	LVGQPSPAYFWSKEGQQLLFPSPVSADGR	KVSPGTGLTIEEVR	384
Qy	384	QVDEGAYVCAGMNS	SAGSSLSKAALKATFETKGRVQKKKSKMGKQKQ	KNVQSIILYISAV	443
Db	385	QVDEGAYVCAGMNS	SAGSSLSKAALKVT--TK-----	-----AV	415

Qy 444 TGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRDGLPIDITDSRISQH 503
Db 416 TGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRDGLPIDITDSRISQH 475
Qy 504 STGSLHIADLKPPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPT 563
Db 476 STGSLHIADLKPPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPT 535
Qy 564 QPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLGQTFNIPDYVASTEYRIKGL 623
Db 536 QPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLGQTFNIPDYVASTEYRIKGL 595
Qy 624 KPSSHYMFVIRAEKNGIGTSPVSSALVTTSPKAAQVALSDKNKMDMAIAEKRLTSEQLI 683
Db 596 KPSSHYMFVIRAEKNGIGTSPVSSALVTTSPKAAQVALSDKNKMDMAIAEKRLTSEQLI 655
Qy 684 KLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTNDNQYVNVTSPTENYVVSNL 743
Db 656 KLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTNDNQYVNVTSPTENYVVSNL 715
Qy 744 MPFTNYEFFVPIYHSGVHSIHGAPSNMMDVLTAEAPPSLPEDVIRMLNLTTLRISWKA 803
Db 716 MPFTNYEFFVPIYHSGVHSIHGAPSNMMDVLTAEAPPSLPEDVIRMLNLTTLRISWKA 775
Qy 804 PKADGINGILKGFQIVIGQAPNNRNNTTNERAASVTLFHLVTGMTYKIRVAARSNGGV 863
Db 776 PKADGINGILKGFQIVIGQAPNNRNNTTNERAASVTLFHLVTGMTYKIRVAARSNGGV 835
Qy 864 GVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVPVIVIVAILIIFVVIAYCY 923
Db 836 GVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVPVIVIVAILIIFVVIAYCY 895
Qy 924 WRNSRNSDGKDRSFIKINDGSVHMASNNLWDVAQNPNQNPMTYTAGRTMNNRNGQALYS 983
Db 896 WRNSRNSDGKDRSFIKINDGSVHMASNNLWDVAQNPNQNPMTYTAGRTMNNRNGQALYS 955
Qy 984 LTPNAQDFNNCDDYSGTMHRPGSEHHYHQAQLTGGPGNAMSTFYGNQYHDDPSPYATT 1043
Db 956 LTPNAQDFNNCDDYSGTMHRPGSEHHYHQAQLTGGPGNAMSTFYGNQYHDDPSPYATT 1015
Qy 1044 LVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADHTAGRRSRSRASDGRGLNGGLH 1103
Db 1016 LVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADHTAGRRSRSRASDGRGLNGGLH 1075
Qy 1104 HRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTGERTPPNKTLMDFIPPPSNPPPP 1163
Db 1076 HRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTGERTPPNKTLMDFIPPPSNPPPP 1135
Qy 1164 GGHVYD---TATRRQLNRGSTPREDTYSVSDGAFARVDVNARPTSRNRNLGGRPLK 1219
Db 1136 GGHVYDDIFQTATRRQLNRGSTPREDTYSVSDGAFARVDVNARPTSRNRNLGGRPLK 1195
Qy 1220 RDDDSQRSSLMDDGGSSEADGENSEGDVPRGGVRKAVPRMGISASTLAHSCYGTNGTA 1279
Db 1196 RDDDSQRSSLMDDGGSSEADGENSEGDVPRGGVRKAVPRMGISASTLAHSCYGTNGTA 1255
Qy 1280 QRFRSIPRNGIIVTQEQT 1297
Db 1256 QRFRSIPRNGIIVTQEQT 1273

RESULT 2
T29548
hypothetical protein ZK377.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29548
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK377.
A:Reference number: Z20639
A:Accession: T29548

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-874 <NHA>
A:Cross-references: EMBL:U88183; PIDN:AAB52657.1; GSPDB:GN00028; CESP:ZK377.2
A:Experimental source: strain Bristol N2; clone ZK377
C:Genetics:
A:Gene: CESP:ZK377.2
A:Map position: X
A:Introns: 91/2; 356/1; 452/1; 701/3; 746/3; 850/1

Query Match 67.5%; Score 4628; DB 2; Length 874;
Best Local Similarity 100.0%; Pred. No. 4.5e-245;
Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 MGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPG 483
Db 1 MGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPG 60
Qy 484 ISWLRDGLPIDITDSRISQHSSTGSLHIADLKPPDTGVYTCIAKNEDGESTWSASLTVEDH 543
Db 61 ISWLRDGLPIDITDSRISQHSSTGSLHIADLKPPDTGVYTCIAKNEDGESTWSASLTVEDH 120
Qy 544 TSNAQFVRMPDPSNFPSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLG 603
Db 121 TSNAQFVRMPDPSNFPSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLG 180
Qy 604 QTFWNIPDYVASTEYRIKGLKPSHSYMFVIRAEKNGIGTSPVSSALVTTSPKAAQVALS 663
Db 181 QTFWNIPDYVASTEYRIKGLKPSHSYMFVIRAEKNGIGTSPVSSALVTTSPKAAQVALS 240
Qy 664 DKNKMDMAIAEKRLTSEQLIKLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTN 723
Db 241 DKNKMDMAIAEKRLTSEQLIKLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTN 300
Qy 724 DNQYVNVTSPTENYVVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNMMDVLTAEAPPSLP 783
Db 301 DNQYVNVTSPTENYVVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNMMDVLTAEAPPSLP 360
Qy 784 PEDVIRMLNLTTLRISWKAPKADGINGILKGFQIVIGQAPNNRNNTTNERAASVTLF 843
Db 361 PEDVIRMLNLTTLRISWKAPKADGINGILKGFQIVIGQAPNNRNNTTNERAASVTLF 420
Qy 844 HLTGMTYKIRVAARSNGGVGVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVP 903
Db 421 HLTGMTYKIRVAARSNGGVGVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVP 480
Qy 904 VIVIVAILIIFVVIAYCYWRNSRNSDGKDRSFIKINDGSVHMASNNLWDVAQNPNQNP 963
Db 481 VIVIVAILIIFVVIAYCYWRNSRNSDGKDRSFIKINDGSVHMASNNLWDVAQNPNQNP 540
Qy 964 MYNTAGRTMNNRNGQALYSLTNAQDFNNCDDYSGTMHRPGSEHHYHQAQLTGGPGNA 1023
Db 541 MYNTAGRTMNNRNGQALYSLTNAQDFNNCDDYSGTMHRPGSEHHYHQAQLTGGPGNA 600
Qy 1024 MSTFYGNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADHTAG 1083
Db 601 MSTFYGNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADHTAG 660
Qy 1084 RRSRSRASDGRGTLNGLLHRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTGERRT 1143
Db 661 RRSRSRASDGRGTLNGLLHRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTGERRT 720
Qy 1144 PPNKTLMDFIPPPSNPPPGGHVYDTATRRQLNRGSTPREDTYSVSDGAFARVDVNAR 1203
Db 721 PPNKTLMDFIPPPSNPPPGGHVYDTATRRQLNRGSTPREDTYSVSDGAFARVDVNAR 780
Qy 1204 PPSRNRNLGGRPLKGRDDDSQRSSLMDDGGSSEADGENSEGDVPRGGVRKAVPRMGI 1263
Db 781 PPSRNRNLGGRPLKGRDDDSQRSSLMDDGGSSEADGENSEGDVPRGGVRKAVPRMGI 840
Qy 1264 SASTLAHSCYGTNGTAQRFRSIPRNGIIVTQEQT 1297

Db 841 SASTLAHSCYGTNGTAQRFRSIPRNNIGVTQEQT 874

RESULT 3

T29549

hypothetical protein ZK377.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T29549

R:Nhan, M.; Hawkins, J.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid ZK377.

A:Reference number: Z20639

A:Accession: T29549

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-423 <NHA>

Cross-references: EMBL:U88183; PIDN:AAB52658.1; GSPDB:GN00028; CESP:ZK377.3

Experimental source: strain Bristol N2; clone ZK377

Genetics:

A:Gene: CESP:ZK377.3

A:Map position: X

A:Introns: 24/1; 142/3; 229/3; 284/2; 408/3

Query Match 32.5%; Score 2232; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-114;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYLGFYHTHTHTYINFDKIPNASLAPVIEHPIDVVSRGSPATLNCGAKPSTAKI 60

Db 1 MYLGFYHTHTHTYINFDKIPNASLAPVIEHPIDVVSRGSPATLNCGAKPSTAKI 60

Qy 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKLVNSGKNGKDSAGAYVCVSNHEGEVKS 120

Db 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKLVNSGKNGKDSAGAYVCVSNHEGEVKS 120

Qy 121 NEGSLKLAHLREDFRVRPRTVQALGEMAVLECSPPRGFPPEPVVSWRKDDKELRIQDMR 180

Db 121 NEGSLKLAHLREDFRVRPRTVQALGEMAVLECSPPRGFPPEPVVSWRKDDKELRIQDMR 180

Qy 181 YTLHSDGNLIIDPDRSDSGTYQCVANMVGERSVNPAPLSVFEPKPEQEPKDMTVDVG 240

Db 181 YTLHSDGNLIIDPDRSDSGTYQCVANMVGERSVNPAPLSVFEPKPEQEPKDMTVDVG 240

Qy 241 AAVLFDCRVGTGDPQPIITWKRKNPEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300

Db 241 AAVLFDCRVGTGDPQPIITWKRKNPEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300

Qy 301 AGTLESAHLRVQAPPSFQTKPADQSVAPAGGTATFECTLVGQPSPAYFWSKEGQDLFP 360

Db 301 AGTLESAHLRVQAPPSFQTKPADQSVAPAGGTATFECTLVGQPSPAYFWSKEGQDLFP 360

Qy 361 SYVSADGRTKVSPTGLTIEEVQRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQK 420

Db 361 SYVSADGRTKVSPTGLTIEEVQRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQK 420

Qy 421 KSK 423

Db 421 KSK 423

Qy 421 KSK 423

Db 421 KSK 423

RESULT 4

T30805

dutt1 protein - mouse

N:Alternate names: transmembrane receptor protein Robol homolog

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30805

R:Wu, M.C.; Lowe, N.; Fordham, R.; Rabbits, P.

submitted to the EMBL Data Library, July 1998

A:Description: The mouse homologue of human DUTTL/H-robol gene: protein sequence and chr

A:Reference number: Z20879

A:Accession: T30805

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1612 <WUM>

A:Cross-references: EMBL:Y1793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1

A:Experimental source: brain

C:Genetics:

A:Gene: dutt1

A:Map position: 16

Query Match 21.9%; Score 1505.5; DB 2; Length 1612;
Best Local Similarity 32.1%; Pred. No. 3.3e-74;
Matches 402; Conservative 176; Mismatches 498; Indels 175; Gaps 34;

Qy 30 PVIEHPIDVVSRGSPATLNCGAK-PSTAKITWYKDGQPVITNKEQVNSHRIVLDTGSL 88

Db 29 PRIVEHPSDLIVSKGEPATLNCAGRPTPTIEWYKGERVETDKDPRSHRLPLSGSL 88

Qy 89 FLKLVNSGKNGKDSAGAYVCVSNHEGEVKSNEGSLKLAHLREDFRVRPRTVQALGEM 148

Db 89 FFLRIVHGRKSR-PDEGVYICVARNYLGEAVSHNASLEVAIRDDFRQNPQSDVMVAVGEP 147

Qy 149 AVLECSPPRGFPPEPVVSWRKDDKELRIQDMRYTLHSDGNLIIDPDRSDSGTYQCVANN 208

Db 148 AVMECQPPRGHPPTISWKKDGSPLDDK-ERITIRG-GKLMITYTRKSDAGKYVCVGN 205

Qy 209 MVGERVSNAPLSVFEPKPEQEPKDMTVDVGAALFDCRVGTGDPQPIITWKRKNPEMPV 268

Db 206 MVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDGELPK 265

Qy 269 TRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLESAHLRVQAPPSFQTKPADQSV 328

Db 266 SR-YEIRDDHLKIRKVTAGDMGSGYTCVAENMVGAESAATLTVQEPHVVKPRDQVVA 324

Qy 329 AGGTATFECTLVGQPSPAYFWSKEGQDLFPSPY--VSADGRTKVSPTGLTIEEVQRQV 386

Db 325 LGRVTFQCEATGNPQPAIFWRREGSQNLFF-SYQPPQSSSRFSVSQGLDITITNVQRSD 383

Qy 387 EGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKQNVQSIKILISAVTGN 446

Db 384 VGYIICQTLNVAGSIITKAYLE-----VTDV 409

Qy 447 TPAKPPPTIEHQNQTLNMGSSAILPCQASGKPTGISWLRDGLPIDITDSRISQHS 506

Db 410 IADRPPPVIRQGPVNVQTVAVDGLTILSCVATGSPAPTLIRKDGVLVSTQDSRIKQLESG 469

Qy 507 SLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSPNPSP 566

Db 470 VLQIRYAKLGDGTGRYCTASTPSGEATWSAYIEVQEGFVGPVPPRPTDNLIPSAPSKPE 529

Qy 567 IVNVTDTEVELHWNAPSTSGAGPITGYIIQYSPDLGQTFNIPDYVASTEYRIKGLKPS 626

Db 530 VTDVSKNVTLNWSQPNLNSGATP-TSYIIIEAFSHASGSWQTAENVTETETFAIKGLKPN 588

Qy 627 HSYMFVIRAEENERKIGTPS-VSSALVTTSKPAQVALSDKNKMDMAIAEKRLTSEQLIKL 685

Db 589 AIYLFVRAANAYGSDPSQISDPVKTQDVPTSGVGHKQ-----VQRELGNVVLHL 641

Qy 686 EEVKTINSTAVRLFVKKRKLLELIDGYIKWRGPPRTN-DNQYV--NVTSPSTENYVVS 742

Db 642 HNPTILSSSSVEYHWTVDQQSQYIYQYKILYRPSGASHGESEWLVEFVRPTPKNSV 701

Qy 743 LMPFTNYEFFVPIYHSGVHSIHGAPSNMVDVLAEPSPSPEDVRIRML--NLITLRIS 800

Db 702 LRKGVNYEIKARPF---PNEFQADSEIKFAKTLLEAPSAAPRSVTVSKNDGNGTALVT 758

Qy 801 WKAPKADGINKLGFQIVIVGQAPNNNNITTNERAASVTLFHLVGTMTYKIRVAARN 860

Db 759 WQPPPEDTQNGMVQYKVMCLNETKYHINKTVDGSTFSVVPISLVPVGIYSVEVAAS 818

Qy 861 GGVGV-----SHGTSEVIMNQDLEKHLAAQENESFLYGLINKSHVPVIVIVAI 910

Db 861 GGVGV-----SHGTSEVIMNQDLEKHLAAQENESFLYGLINKSHVPVIVIVAI 910

```

10  PVIIEHPIDVVVRSRGPATLNCAGK-PSTAKITWYKDGQPVITNKEQVNSHRVLDTGSL 88
   ||:|:| |:|:|:| ||||| |: | ||| |: |:|: |||:|:|:|
Db 68  PRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHMLLP6GSL 127

Qy 89  FLLKVNKSGKNGKSDAGAYIVCVASNEHGEVKSNEGSLKMLAREDFVRPRVTQALGGM 148
   ||:|:| |:|:|:| ||||| |: | ||| |: |:|:|:|:|:| | | ||
Db 128 FFLRIYHVGKRSR-PDEGVYICVARNYLGEAVSHNASLEVALLRDDFRQNPSDVMVAVGEP 186

Qy 149 AVLECSPPRGFPEPVVSWRKDDKELRIQDMPRYTLHSDGNLLIDPVDRSDSGTYQCVANN 208
   ||:|:| |||| |:|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 187 AVMECQPPRGHPEPTISWKGDSPLDDKD-ERITIRG-GLMITYYTRKSDAGKYVCVGN 244

Qy 209 MVGERVSNPARLSVFEPKPFQEPKDMTVDVGAAVLFCRVTDGDPQITWKRKNPEMPV 268
   ||||| |:|:|:| |:|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 245 MVGERESKVDVTVLERPFSFYKRPSNLAVTVDDSAEFKCEARGDPVPTFGWRKDDGELPK 304

Qy 269 TRAYIAKDNRLRIERVQPSDEGEVYCIARNPAGTLEASAHLRVQAPPSFQTKPADQSV 328
   |:| |:|:| |:|:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|
Db 305 SR-YEIRDHTLKIRKVTAGDMGSYTCVAENMVGKAEASATLTQVEPPEPHVVKPRDQVVA 363

Qy 329 AGGTATFECTLVGPSPAPFWSKEGQDQLLFPST--VSADGRTKVSPTGTLTIEEVQRVD 386
   |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:| |:|:|

```

```

RESULT 6
T14316
rig-1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14316
R;Yuan, S.S.F.; Cox, L.A.; Dasika, G.K.; Lee, E.Y.H.P.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z17975
A;Accession: T14316

```


Qy 36 PIDVVVSRGSPATLNCGAKPST-AKITWYKDGQPVITNKEQVNSHRIVLDTGSLFLKVN 94
|:|:: ||: :|| : | || | ||| : | : | :| ||| : |

Db 25 PMDILSVRGASVIMCNSYCYETPPKIEWKKDG--TLLNLVS--DDRRQLPLDGSLLINSV 81

Qy 95 SGKNGKDSAGAYYCVASNEH-GEVKSNEGSLKAMLRDFVRPRPTVQALGGEMAVLEC 153
| : | | | | : | : | : | : | : | : | : | : |

Db 82 HSKHNK-PDEGYQCVAIVESLSIVSRITAKTVAGLPR-FTSQPELSVYKGSAILNC 139

Qy 154 SPPRGFPPEPVVSWRKDDKELRIQDMPRYTLHSDGNLIIDPVDRSDSGTYQCVANNMVGER 213
| : | : | : | : | : | : | : | : | : | : | : |

Db 140 EVNVDL-APFVRWEQDRQPLSLDD--RVFKLPSCALLIGNATDDGGFYRCVIESGGTPK 196

Qy 214 VSNPARLSVPEKPK-----FEQEPKDMTVDVGAALFDCRVTDGDPQOPITWKRRNEPMP 267
| : | : | : | : | : | : | : | : | : | : | : |

Db 197 YSEAEELKILPDPEEPQSLFVVRQPSLLIKTVTQNAVFPVCVAGGFPTPYVRWTKNGEELI 256

Qy 268 V--TRAYIAKNRGLRIERVQPSDEGEYVCYARNPAGTLEASALRVQAPPSFQTKPADQ 325
: : : : : | : | : | : | : | : | : | : |

Db 257 TEDSERFALRAGGSLISDVTEEDVGTYTICADNENETIEAQELAVQVPPEFLKRPANI 316

Qy 326 SVPAGGTATFECTLVGQSPPAYFWSKEGQQDLFFSPYSVADGRTKVSPTGTLTIEEVRQV 385
| | : | : | : | : | : | : | : | : | : |

Db 317 YAHESMDIVFECEVTGKPTPTVKWVNG--DVVIPS-----DYFKIVKEHNLQVLGLVKS 369

Qy 386 DEGAYVCAGMNSAGSSLSKAAL-----KAT----- 410
| | | | : | : | : | : | : |

Db 370 DEGPHYCIAENDVGNAQAGALIILDLDAIPTLPTSLTSATNDHLAPATTGPLPTAPR 429

Qy 411 -----FETKGRVQKKK----SKMGKQKQKNVQS 434
| | : : : | : | : | : |

Db 430 DVVATLVSTRFIRLTWRTPVSPDQGNLTYSIFYTKGINRREVENTSRPG-ETQVMIQN 488

Qy 435 IIK---YLISAVTGNPAKPPPTIEHGRQNTLMVGSSSAILPCQASGKPTPGISWLDRGL 491
: : | : | : | : | : | : | : | : | : |

Db 489 LMPETVYVFRVVAQN-----KHGHESSAPLKVATQPEVQLPG-PAPNIR-AYAGS 537

Qy 492 PIDIT---DSRIS-----QHSTGSLHIADLKKPDTGV-----YTCIA 525
| : | : : | : | : | : | : | : | : |

Db 538 PTSVTVTWETPLSGNGEIQNYKLYMEKGQDSEQDQVDVAGLSYITGLKKYTEYSFRVVA 597

Qy 526 KNEDGESTWSASLTVEDHTSNAQFVRMPDPNSPFPSSPTQPIIVNVTDTE-VELHWN-APS 583
| : | : | : | : | : | : | : | : | : |

Db 598 YNKHGPGV-----STQDVVVRTLSDPVSAAPQLNLTLEARNKSLMLHWQPPPA 645

Qy 584 TSGAGPITGYIIQY----- 597
: : | | | : |

Db 646 GTHSGQITGYKIRYRKVSRSKSDVTESVGGTQLFQLIEGLERGTEYNFRIAAMTVNGTGPA 705

Qy 598 -----YSPDLGQTWFENIPDYASTE----- 617
: | | : : | : |

Db 706 TDWVSAETPESDLDES--RVPEVPSLSLHVRPLVTSIVVSWTPPENQNIIVRGYAIGYGIG 763

Qy 618 -----YRIKGLKPSHSYMFVIRAENEKGIGTSPVSALLVTSKPAQVAL 662
| : | : | | : | : | : | : | : | : |

Db 764 SPHAQTIKVDYKQRYTTIENLDPSHHYVITLKAFNNVGEIPLYESAV---TRPH----- 815

Qy 663 SDKNKMMAIAEKRLT-----SEQLIKL-EEVKTINSTAVRLFWRKRKL-----EELIDG-- 711
| | : : : | : | : | : | : | : | : | : |

Db 816 SDTSEVDLFEVFNAPYTPVPDPSPMPVPGVQASILLSHDTIRITWADNSLPKNQKITDARY 875

Qy 712 YYIKWRGPPTFN--DNQYVNVTSPTSTENYVSNLMPFTNYEFFVI---PYHSGVHSI--H 764
| : : | : | : | : | : | : | : | : | : |

Db 876 YTVRW----KTNIPANTKYKTANATTLSTLYTGLKPNLTYEPVSMVTKGRSSTWSMTAH 931

Qy 765 GAPSNSMDVLTAEPAPSLPPEDVRI--RMLNLTLLRISWKAPKADGINGILKGFQIVIVG 822
| | : | : | : | : | : | : | : | : |

Db 932 GT-----TFELVPTSPKDVTVVSKEGKPTIIVNWQPPSE--ANGKITGY-IIYYS 980

Qy 823 QAPNNNRNITINERAASVTLFHLVTGMT----YKIRVAARSNGGVG----- 854
| : | : | : | : | : | : | : | : |

Db 981 TDVNAEIHWDVIEPVVGNRLTHQIQELTLDTPYYFKIQARNSKGMGPMSAVQFRTPKAE 1040

Qy 865 -----VSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVPVIVIVAIL 911
| : | : | : | : | : | : | : | : |

Db 1041 SSDKMPNDQASGAGSKSRPVDVGDYKPKPLSGNSPHGSPSTSPIDSNMLLVIIIVSGVVI 1100

```

Qy      30 PVIIIEHP---IDVVSRGSPATLNCGAKPS-TAKITWYKDGQPVTINKEQVNSHRIVLTD 85
      || : | | | | ||| : : : | : | : : : || : |
Db      26 PVFIKEPNSIIEVDSEDKKITLNCCEARGNPSPHYRWQLNSDDITSLD---HRYKLNG 81
      || : | | | | ||| : : : | : | : : : || : |
Qy      86 GSLFLLKVSNGKNGKSDAGAYCYVASNEHEGVKSNESGLKLAMLREDFRVRPR-TVQAL 144
      | : | : | | : : | : | | | | : | : | | : | : | |
Db      82 GNLIIVNPN-----RWNDTGSYQCFATNSLTIVSREAKLQFAYL-ENFKTRMRSTVSVR 135
      || : | | | | ||| : : : | : | : : : || : |
Qy     145 GGEMAVLECSPPRGFPEPVVSWRKDDKELRI-QDMPRYTLHSDGNLIIDPVDRSDSGTYQ 203
      | : | | | | : : | : : : | : | : | | | |
Db     136 EGQGVVLLCGPPPHSGELSYAWWFNEYPSFVEEDSRRLVSQETGHLHYIAKVEPSDVGNWT 195
      || : | | | | ||| : : : | : | : : : || : |
Qy     204 CVANNMVGER--VSNPARLSVFE-----KPKFE-QEPKDMTVDGAAVLFDCRVTDGP 253
      || : | : : | : | : : | : | : : : || : |
Db     196 CVWTSTVTNTIRVLGSPPTPLVLRSDGVMEYEPKIEVQFPETLPAAGKSTVRLECFALGNP 255
      || : | | | | ||| : : : | : | : : : || : |
Qy     254 PQQITWKRKN-EPMPVTRAYIAKDNRGLRIERVQPSDEGEYCYVARNPAGTLEASAHLRV 312
      || | | | | | | | | | | | | | | | | | |
Db     256 VPOINWRRSDGMPPP-NKIKLRKFNGLRIENFOOEDTGSYEGIAENSRGKNVARGRLTY 314
      || | | | | | | | | | | | | | | | | | |

```

```

Qy 313 QAPPSFQTKPADQSVPAAGTATFECTLVGQSPSPAYFWSKEGQDQLLFPSYVSADGRTKYVS 410
   ||: | : : || :||:| | | :| :| :|
Db 315 YAKPYWLQLLRDVEIAVEDSLYWECRASGPKPKPSYRWLKNGLDALVL-----EERIQIE 367

Qy 373 PTGTLTIEEVRQVDEGAYVCAGMNSAGSSLSKAALKAT----- 410
   ||| : | : | | | | | |
Db 368 -NGALTITNLNLTDSGMFQCIENKHGLIYSAAELKVVASAPDFSRRNPKMKMVQVQVGS 426

Qy 411 --FETKGRVQ-----KKSKMGKQKQK-----NVQSIK-----YLISAV---- 443
   : | | : | : : : : : : : : : :
Db 427 VILDCKPRASPRALSFWKKGDMVMREQARVSFLNDGGLKIMNVTKADAGYITCTAENQFG 486

Qy 444 ----TGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGI----SWLRDGLPIDITD 497
   | : | : | : | | | | | : | : |
Db 487 KANGTTHLVVIEPTRIILAPSNMDVAVGESVILPCQVQVHDLPLDIMFAWYFNGALTDFFK 546

Qy 498 ----SRISQHSTGSLHIADLKPDVTGYTICAKNEDGESTWSASLTVEDHTSNAQFVRM 552
   : : | : | : : : | : : : : :
Db 547 DGSHFKEVGGSSGDLMRNIQLKHSKRVCMQVTGVDVSAAELIVR----- 595

Qy 553 PDPSNFPSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYYS-PDLG-QIWFNIP 610
   | | : : | : | : | : | : | : | :
Db 596 ---GSPGPPENVKVDEITDTTAQLSW-TEGTDHSHSPVISYAVQARTPFSYVGWQSVRTVP 650

Qy 611 DIVASTEY--RIKGLKPSHSMYFVIRAENEKIGTSPVSSALVTTSKPAQVALSDKNKM 668
   : : | : | : | : | : | : | : :
Db 651 EVIDGKTHATVVELNPMWEYEFRIVASNKIGGGEPSLPSEKVRTEEAAPAIAPSEVS-- 708

Qy 669 DMAIAEKRLTSEQLIKLEEKVTINSTAVRLFWRKKLEELID---GYIKWRGPPPTND 724
   || : : : | | | : | : | : |
Db 709 ---GGGGSRSELVITWDPVP-----EELQNGGGGFYVAFRPLGVTTW 748

Qy 725 NQYVNVTPSPSTENYVVS--LMPFTNYEFFVIPYHSGVHSIHGAPNSMDVLTAEPSSL 782
   | | | | : : | : | : | : | : | :
Db 749 IQTV-VTSPDNPRVFRNESIVFPSEYKVGVIN---NKGEGPFSPTVTVFSAEIEPTV 804

Qy 783 PPEDVRIRMLNLTTLRISWKAPKADGINGILKGFIQVI--VGQAPNNNRNITTNERAASV 840
   : : | : : | | : : : | : : | :
Db 805 APSHISARSLSSSEIEVSNWNTIPWKLNGHLLGEYVRTWNGGEESSRKVKVAGNQTS 864

Qy 841 TLFHLVTGMTYKIRVAARSNGGVGVSHT-----SEVIMNQD 877
   | | : : | : | : | : | : | :
Db 865 VLRGLKSNLAYITAVRAYNSAGAGPFSATVNNATTKTPSPQPPGNVVVNATDTKVLLNWE 924

Qy 878 TLEKHLAAQQENESFLYG 895
   : : | | | : |
Db 925 QVK-----AMENESEVTG 937

```

```

RESULT 10
T13822
frazzled gene protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13822
R;Kolodziej, P.A.; Timpe, L.; Mitchell, K.J.; Goodman, C.S.; Fried, S.; Jan, L.Y.; Jan
Cell 87, 197-204, 1996
A;Title: Frazzled encodes a Drosophila member of the DCC immunoglobulin subfamily and
A;Reference number: z17780; MUID:97015076
A;Accession: T13822
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1375 <XOL>
A;Cross-references: EMBL:U71001; NID:g1621114; PID:g1621115; PIDN:AAC47314.1
C;Genetics:
A;Gene: frazzled
A;Map position: 2
C;Function:
A;Description: may function in vivo as a receptor or component of a receptor mediati

```

Query Match 8.8%; Score 607; DB 2; Length 1375;
Best Local Similarity 21.6%; Pred. No. 3e-25;

```

Qy      16 YINFDKIPNASNLAPVIIIEHP-----DVVVSGRSGPATLNCGAKPS-TAKITWYKDG 66
      :: ||: |:| :|:| :||| :|:| | : ||| |
Db      372 FVRKDKL-SAQDYQVVLDEGTGPKIIISAFSEKVVSPAEPVSLMNCVKGTPLPTITWILDD 430

Qy      67 QPVITNKEQVNSHRI---VLDTGSLF-LLKVSNGKNGKDSAGAYCYVASNEHEGVKSNE 122
      |:| :| ||| : |:| :|:| : ||| | |:| | |
Db      431 DPILKG---GSHRISQMITSEGNVVSYLNISS---SQVRDGGVYRCTANNSAGV---- 479

Qy      123 GSLKLAMLREDFR---VRP-RTVQALGGEMAVLECSPPRGFPPEPVSVWRKDDKELRIQD 177
      | |:| :| |:| : |:| :| |:| :| |:| |
Db      480 ---LYQARINVRGPASIRPMKNITAIAGRDYIHC-RVIGYPYYSIKWYKNSNLLPFNH 534

Qy      178 MPRYTLHSDGNLIIDPVDRS-DSGTYQCVANNMVGERVSN-----PARLSVFKEP 226
      : :| |:| :| :| |:| | |:| :| :| |:| |
Db      535 R-QVAFENNGTKLSDVQKEVDEGEYTC--NVLVQPQLSTSQSVHVTVKVPFPIQPEFP 591

Qy      227 KFEQEPKMDTVDGAAVLEDC-RVTGDPOPOITWKRKNEMPMVTRAYIAKDN----RGLR 281

```

Db 592 RF-----SIGQRVFIPCVVSGDLPTITWQKDRPIPGSLG-VTIDNIDFTSLR 641
Qy 282 IERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPPSTQKPADQSVAGTATFECTLVG 341
Db 642 ISNLSLMHNGNYTCIARNEAAVEHQSQLIVRVPPKFFVQPRDQDGIYKAVILNCSAEG 701
Qy 342 QPSPAYFW--SKEGQDQLFPSTVSADGRKTVSPGTGLTIEEVRQVDEGAYVCAGMSAG 399
Db 702 YPVPTIVWKFSGAGVPQFPQ--IALNGRIQVLSNGSLLIKHVVEEDSGYILCKVSNVVG 759
Qy 400 SSLSKAALKATPFTKGRVQKKSKMGKQKQKQVQSIILYISAVTGNTPAKPPPTIEHGH 459
Db 760 ADVSKS-----MYLTVKI-----PAMITSY 779
Qy 460 QNQTLMV-GSSAILPCQASGKPTPGISWLRDGLPID-----ITDSRISQHSSTGLSHIA 511
Db 780 PNTTLATQGGKEMSCTAHGEKPIIVRWEKEDIINPEMAYLVSTKEVGEVISTLQIL 839
Qy 512 DLKKPDTGVYTCIAKNEDGESTWSASLIVEDTSNAQFVRMPDPNFPSSPTQPIIVNVT 571
Db 840 PTVREDSGFFSCHAINS YGEDRGIIQLTVQE-----PPDPPEIEIKDKV 883
Qy 572 DTEVELHWNAPSTSGAGPITGIIQYSPDLGQTFWNI-----PDYVASTERYIKGLK 624
Db 884 ARTITLRWTM-GPDGNSPITGYDIE--CKNKSWSDSAQRTKDVSPQLNSAT---IIDIH 937
Qy 625 PSHSYMFIARAENEGIGITPSVSALVTTSKPAAQVALSDKNKMDMAIAEKRLTSEQLIK 684
Db 938 PSSYSIRMYAKNR--IGKSEPSNELTITADEAAPDG-----PPQEVH 978
Qy 685 LEEVKTIINSTAVRLFVK--KRKLEE-LIDGYIKWRGPPRTNDNQYVNV---TSPSTEN 737
Db 979 LE---PISSQSIRVTWAPKKHLQNGIIRYQIGYR-EYSTGGNQFNIIISVDPSGDSEV 1034
Qy 738 YVVSNLMPFTNYEFFVPIPHSGVSHGAPSNSMDVLTA--EAPPSLPPEDVRIRMLNLT 795
Db 1035 YTLONLNFQYGLVQACNRA----GTGPPSSQEIITTTLEDVPSPPENVAIATPE 1089
Qy 796 TLRISWKAPKADGINGILKGFQIVIVGQAPNNN---RNITTNERAASVILFHLVTGMTY 851
Db 1090 SISISWSTLSKEALNGILQGFRIYVWNLMDGELGEIKNITTQ--PSLELDGLEYKTYN 1147
Qy 852 KIRVAARSNGGVGV 865
Db 1148 SIQVLAFTAGDGV 1161

RESULT 12

JU0094

F11 protein precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 31-Jan-2000

C:Accession: JU0094

R:Bruemendorf, T.; Wolff, J.M.; Frank, R.; Rathjen, F.G.

Neuron 2, 1351-1361, 1989

A:Title: Neural cell recognition molecule F11; homology with fibronectin type III and in

A:Reference number: JU0094; MUID:90180453

A:Accession: JU0094

A:Molecule type: mRNA

A:Residues: 1-1010 <BRU>

A:Cross-references: GB:X14877; NID:g1708784; PIDN:CAA33018.1; PID:g63385

A>Note: the carboxy-end hydrophobic stretch is compatible with the consensus motif for C

A>Note: f11 comprises six domains related to the immunoglobulin domain type C and four C

C:Comment: F11 is a chick neural cell surface-associated glycoprotein implicated in neur

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

C:Keywords: blocked carboxyl end; glycoprotein; lipoprotein; phosphatidylinositol linkag

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-984/Product: protein F11 #status predicted <MAT>

F:247-303/Domain: immunoglobulin homology <IMM>

F:985-1010/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F:200,249,329,448,464,485,512,582,621,924/Binding site: carbohydrate (Asn) (covalent) #s

F:984/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature

Query Match 8.7%; Score 597; DB 2; Length 1010;
Best Local Similarity 23.6%; Pred. No. 6.7e-25;
Matches 268; Conservative 140; Mismatches 364; Indels 364; Gaps 47;

Qy 6 FYHTHTHTYII-----NFDKIPNASLAPVIEHPIDVVV---SRGSPATLNCAGKP 55
Db 3 FFISHLVTLCPFCVADSTHFSEEGN-KGYGPVFEEQPIDTIYPEESSDQGVSMNCRAR- 60
Qy 56 STAKITWKDQGPVITNKEQVNSHRIVL--DTGSLFLKVNKNGKNGKSDAGAYVCASN 113
Db 61 -----AVPPPTYKWLNNWDIDLTKDRISMVGGRLVISNPEKSRDAGKYVCVSN 110
Qy 114 EHEGVKSNEGSLKLAHL-----REDFRVRPRVTQALGEMAVLECSPPRGFPPEV-VSWR 167
Db 111 IFGVRSSEATLSFGYLDPPPEEHYEVKVR-----GVGAVLLCEPPHYHDDLSYRWL 165
Qy 168 KDKELRIQ-DMPRYTLHSDGNLIIDPVDRSDSGTYQCVANNMVGERSNPA-RLSVFEK 225
Db 166 LNEFPVFIALORRRFVSQTNGNLIANVEASDKGNYSCE-----VSSPSITKSVSFSK 217
Qy 226 ----PKFEQE---PKDMTVD-----VGAALFDCRVTGDPQPIITWKRKNEMPVT- 269
Db 218 FIPLIPQADRAKVYPADIKVKFKDTYALLQGNVLTLECFALGNVPVLEWSKYLEPMPATA 277
Qy 270 -----RAYI-----AKD--- 276
Db 278 EISMSGAVLKIFNIQYDEGLYECEAENYKGDQKQARVYVQASPEWVEHINDTEKDIGS 337
Qy 277 -----NRG-LRIERVQPSDEGEYVCYARNPAGTLEASAH 310
Db 338 DLYWPCVATGKIPTIRWLKNGVSRFKGELRIQGLTFEDAGMYQCIENAHGIIYANAEL 397
Qy 311 RVQA-PPSFQKPADQSVPA--GGTATFECTLVGQPSPAYFWSKEGQDQLFPSTVSADG 367
Db 398 KIVASPTTFELNPMKKKILAAGGRVITECKPAAPKPKFSWSK-GTELLVNGS----- 450
Qy 368 RTKVSPTGLTIEEVRQVDEGAYVCAGMSAGSSLSKAALKATPFTKGRVQKKSKMGKQ 427
Db 451 RIHWDDGSLIINVTKLDEGRYTCFAENNRKANSTGVLEMTATR----- 497
Qy 428 KQKNVQSIILYISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGIS-- 485
Db 498 -----ITLAPLVNVDVTGENATMQCIASHDPTLDLTFI 530
Qy 486 WLRDGLPIDIT-----DSRISQHSSTGLSHIADLKKPDTGVYTCIAKNEDGESTWSALT 539
Db 531 WSLNGFVIDFEKEHEHYERNVNIKSGELLIKVQLRHAGRYTCTAQTIVDNSSASADLV 590
Qy 540 VEDHTSNAQFVRMPDPNFPSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGIIQYYS 599
Db 591 -----VRGP-----PGPPGRIEIEIRDVAVALTWSR-GTDNHSPISKYTIQ-SK 633
Qy 600 PDLGQTFWNI PDYVASTE-----YRIKGLKPSHYSMFVIRAENEGIGITPSVSSA 649
Db 634 TPLSEEWKD-----AKTEPSDIEGNMESARVIDLIPWMEYEFRIATNLGTGEPSPMSQ 688
Qy 650 LVTISKPAQVALSD-----KNKMDMAIAEK-----RL 677
Db 689 RIRTEGAPPNVAPSDVGGGGGSGNRELITWMPLSREYHYGNFQIVAFKPFGEKEWRRV 748
Qy 678 T-----SEQLIKLE-----EV 688
Db 749 TVTNPEIGRYVHKDESMPPSTQYQVKVAFNSKGDGPFSLTAVIYSAQDAPTEVPTDVSV 808
Qy 689 KTINSTAVRLFVKRLEELIDGYIYK-WRGPPRTNDNQYVNVNTPSTENY--VVSNLMP 745
Db 809 KYLSSEISVSW-HHVTEKSVGEQIRYWAHDKAAARQVQV---SNQSEYTKLENLKP 864
Qy 746 FTNYEFFVPIPHSGVSHGAPSNSMDVLTAEPSPSPPEDVRIRMLNLTTLR-----I 799

Db 865 NTRYHIDVSANFS---AGYPPSPRTIDIITRKAPPSQRP---RI---ISSVRSGRSRII 914
Qy 800 SWKAPKADGINGILKGFQIVIVGQAPNNNRNITTNERAASVTLFHLVTGMTYKIRVAARS 859
Db 915 TWDHVKAMSNEASAVEGYKVLVYRDPGQHEGLFSTGKHTIEVP---VPSDGEYVVEVRAHS 971
Qy 860 NGGVG-----VSHGTSEVIMNQDLEKHLAAQQENESFLYGLINKSHVPIVIVA 909
Db 972 EGGDGEVAQIKISGATAGV-----PTLLGLV---LPALGVLA 1006

RESULT 13

S05944

neuronal cell surface protein F3 precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 21-Jan-2000

C:Accession: S05944

R:Gennarini, G.; Cibelli, G.; Rougon, G.; Mattei, M.G.; Goridis, C.

J. Cell Biol. 109, 775-788, 1989

A:Title: The mouse neuronal cell surface protein F3: a phosphatidylinositol-anchored mem

A:Reference number: S05944; MUID:89340657

A:Accession: S05944

A:Molecule type: mRNA

A:Residues: 1-1020 <GEN>

A:Cross-references: EMBL:X14943; NID:g50937; PIDN:CAA33075.1; PID:g50938

C:Genetics:

A:Map position: 15F

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

F;1-20/Domain: signal sequence #status predicted <SIG>

F;256-312/Domain: immunoglobulin homology <IMM>

Query Match 8.7%; Score 595; DB 2; Length 1020;
Best Local Similarity 24.0%; Pred. No. 8.7e-25;
Matches 263; Conservative 127; Mismatches 381; Indels 324; Gaps 43;

Qy 4 LGFYHTH-THHTYINFDKIPNASLAPVIEHPIDVVVSRGS---PATLNCAGKPSTAK 59
Db 19 LGDPTWHRRYGHGVSEEDK----GFGPIFEEQPIINTIPEESLEGKVSINCRARASFPF 73
Qy 60 ITWYK----DGQPVITNKEQVNSHRIVLDTGSLFLKVNKSGKNGKDSAGAYYCVASNEH 115
Db 74 V--YKWRMNGVDVLTN-----DRYSMVGGNLVI-----NNPDQKQDAGIYCLASNNY 120
Qy 116 GEVKSNEGSLKLAHLREDFRVRPR-TVQALGGEMAVLECSPPRGFPEPV-VSWRKDD 173
Db 121 GMRVSTEATLSFGYL-DPPFPEERPEVKVKEGKGMVLLCDPPYHFPDDLRYRWLLNEFPV 179
Qy 174 RI-QDMPRYTLHSDGNLIIDPVDRSDGTCYQVANNMGERVSNPA-RLSVFEK----- 225
Db 180 FITMDKRRFVSQTGNLYIANVESSDRGNYSF-----VSSPSITKSVEFSKIPLIP 231
Qy 226 -PKFEQEP-----KDMTVDVGAALFDCRVTDGPPQITWKRKNEPMPVTRAYIAK 275
Db 232 IPERTTKPYPADIVVQFKDIYTMGQNVLTLECFALGNPVPDIRWRKLEPMPST-AEIST 290
Qy 276 DNRGLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPPSFQTKPADQSPVAGGTATF 335
Db 291 SGAVLKIFNIQLEDEGLYECAENIRGDKHQARIYVQAFPEWVEHINDTEVDIGSDLYW 350
Qy 336 ECTLVGQPSPAYFWSKEGQDQLLFPYSYVSADGRKTVSPGTGLTIEEVQVDEGAYVCAGM 395
Db 351 PCIATKPIPTIRWLKNGY-----SY-----HGELRLRYDTFENAGMYQICAE 394
Qy 396 NSAGSSLSKAALK-----ATPETKGRVQKKSKMGKQKQKQVQSIKYLISAVTGNTPAK 450
Db 395 NAYGSIYANAEKLALAPTF-----MNPMKKK-----ILAA----- 427
Qy 451 PPPTIEHGHQNTLMVGSSAILPCQASGKPTGISWLRDGLPIDITDSRISQHSHTGSLHI 510
Db 428 -----KGRVVIIECKPKAAKPKFSWSK-GTEWLVNSRILLWEDSLEI 471
Qy 511 ADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHT-----SNAQFVRMPD 554

Db 472 NNITRNDGGIYTCFAENNRKANSTGLVITNPTRIILAPINADITVENATMQCAASF 531
Qy 555 PS-----NF----- 558
Db 532 PALDLTFVWSFNGYVIDFNKEITHYQRFNMLDANGELLIRNAQLKHAGRYCTAQTIV 591
Qy 559 -----PSSPTQPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLGQ 604
Db 592 DNSSASADLVVRGPPGPGGLRIEDIRATSVALTWSRGSNDHS-PISKYTIQ-TKTILSD 649
Qy 605 TWFNI---PDYVASTERYIK--LKPSHSYMFVIRAENEKIGITPSVSSALVTSKPAAQ 659
Db 650 DWKDAKTPPIIEGNMESAKAVDLIPMEYEFVRVATNTLTGEPISPNRIKTDGAAPN 709
Qy 660 VALSD-----KNKMDMAIAEKRLTSEQLIKLE----- 686
Db 710 VAPSDVGGGGTNRLETTITWAPLSREYHYGNFGYIVAFKPFDEWKKVTVNPDTRY 769
Qy 687 -----EVKINSTAVRL 698
Db 770 VHKDETMTPTAFQVKVAFNKNKGDPYSLVAVINSAQDAPSEAPTEGVKVLSSSEISV 829
Qy 699 FWKRRKLEELIDGYIK-WRGPPTNDNQYVNVTSPTENYV-VSNLMPFTNYEFFVIP 755
Db 830 HW-KHVEKIVESYQIRYAGHDKAAHVRVQTS---QEYSARLENLLPDQYFIEVGA 885
Qy 756 YHSGVHSIHGAPSNMMDVLTAEAPPSPDPDVRIRMLNLTTLR-----ISWKAPKADGI 809
Db 886 CNS---ACGPPSSDIETTRKAPPSQPP---RI---ISSVRSGRSRIITDWHVVALSN 935
Qy 810 NGILKGFQIVIVGQAPNNNRNITTNERAASVTLFHLVTGMTYKIRVAARSNGGVGVSHGT 869
Db 936 ESTVTGYKILYRDPGQHDGKLFSTHKHSIEVP---IPRDEYVVEVRAHSDGGDGV---V 989
Qy 870 SEV-IMNQDLEKHL 883
Db 990 SQVKISGVSTLSSSL 1004

RESULT 14

A57112

contactin precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jan-2000

C:Accession: A57112

R;Peles, E.; Nativ, M.; Campbell, P.L.; Sakurai, T.; Martinez, R.; Lev, S.; Clary, D.
Cell 82, 251-260, 1995

A:Title: The carbonic anhydrase domain of receptor tyrosine phosphatase beta is a fun

A:Reference number: A57112; MUID:95354206

A:Accession: A57112

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: mRNA

A:Residues: 1-1021 <PEL>

C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homolo

C:Keywords: membrane protein; phosphatidylinositol linkage

F;256-312/Domain: immunoglobulin homology <IMM>

Query Match 8.7%; Score 595; DB 2; Length 1021;
Best Local Similarity 23.8%; Pred. No. 8.7e-25;
Matches 257; Conservative 125; Mismatches 371; Indels 326; Gaps 41;

Qy 4 LGFYHTH-THHTYINFDKIPNASLAPVIEHPIDVVVSRGS---PATLNCAGKPSTAK 59
Db 19 LGFTWHRRYGHGVSEEDK----GFGPIFEEQPIINTIPEESLEGKVSINCRARASFPF 73
Qy 60 ITWYK----DGQPVITNKEQVNSHRIVLDTGSLFLKVNKSGKNGKDSAGAYYCVASNEH 115
Db 74 V--YKWRMNGVDVLTN-----DRYSMVGGNLVI-----NNPDQKQDAGIYCLASNNY 120
Qy 116 GEVKSNEGSLKLAHLREDFRVRPR-TVQALGGEMAVLECSPPRGFPEPV-VSWRKDD 170

```

RESULT 15
S01998
contactin precursor - chicken
N;Alternate names: 130K glycoprotein
C;Species: Gallus gallus (chicken)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 21-Jan-2000
C;Accession: S01998
R;Ranscht, B.; Dours, M.T.
J. Cell Biol. 107, 1561-1573, 1988
A;Title: Sequence of contactin, a 130-kD glycoprotein concentrated in areas of interneuron

```

```

Qy      6 FYHTHTHTYI-----NFDKIPNASNLAPVIEHPIDVV---SRGSPATLNCAGAK 55
       |:|:| |:| |:|:| |:| |:| |:| |:| |:| |:|
Db     3 FFISHLVTLCFIFCVADSTHFSEEGN-KGYGPVFEEQPIDIITYPEBSSDQGVSMNCRAR- 60

Qy    56 STAKITWYKDGPVITNKEQVNSHRIVL---DTGSFLFLKVNSGKNKGDS DAGAYICVASN 113
       | :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    61 -----AVFPPTYKWKLNNWDIDLTKDRYSMVGGRLVISNPEKSRDAGKYVCVVSN 110

Qy   114 EHGEVKSNEGSLKLAML----REDFRVRPRTVQALGGEMAVELECSPPRGFPEPV-VSWR 167
       | |:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db   111 IFGTVRSSSEATLSFGYLDPFPPEEHYEKVKRE----GVGAVLLECPYPHYDDLSYRWL 165

Qy   168 KDDKELRIQ-DMPRYTLLHSDGNLIIDPVRSDSGTYQCVANNMVGERSNP-A RLSVFEEK 225
       ::| :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db   166 LNEFPVFIALDRRRFVSQTNGNLYIANVEASDKGNYSCE-----VSSPSITKS VFSK 217

Qy   226 ----PKFEQE--PKDMTVD-----VGAAVLFDCRVTDGPQPIQTKWRKNEMPVT- 269
       |:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db   218 FIPLIPQADRKVPADIKVKFKDYALLGNVLTLECFALGNPNVPELRWSKYLEMPATA 277

Qy   270 -----RAYI-----AKD--- 276
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   278 EISMSGAVLKIFNIQYDEGLYECEAENYKGKKKHQARVYVQASPEWEHINDTEKDIGS 337

Qy   277 -----NRG-LRIERVQPSDEGEYVCYARNPAGTLEASAHL 310
       :| |:|:| :| :| :| :| :| :| :| :| :| :| :|
Db   338 DLYWPCVATGXPIPTIRWLKNGVSFRKGLERIQLTFEDAGMYQCI AENAHHIYANAEL 397

Qy   311 RVQA-PPSFTKPADQSVA--GGTATPECTLVQGQSPAYFWSKEGQQDLLFPSYVSADG 367
       ::| |:|:| :| :| :| :| :| :| :| :| :| :| :| :|
Db   398 KIVASPPTFELNPMKKILAARGGRVIEECKPRAAPKP KFSWSK-GTELLVNGS----- 450

Qy   368 RTKVSPTGLTIIEEVQRVDEGAYVCAGMSAGSSLSKAALKATFETKGRVQKKSKMGQ 427
       | :| |:|:| |:| |:| |:| |:| |:| |:| |:| |:| :|
Db   451 RIHIWDDGLEIINVTKLDEGRYTCAENNRRKANSTGVLEMTEATR----- 497

Qy   428 KQKNVQSIILYKILSAVTGNTPAKPPPTIEHGHQNQNTLMVGSSAILPCQASGKPTPGIS-- 485
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   498 -----ITLAPLNVDVTGENATMQCIASHDPTLDLTFI 530

Qy   486 WLRDGLPIDIT-----DSRISQHS TGSLHIADLKKPD TVYTCIAKNEDGESTWSASLT 539
       | :| |:| :| :| :| :| :| :| :| :| :| :| :| :|
Db   531 WSLNGFVIDFEKEHEHYERNVMIKSN GELLIKNVQLRHAGRYTCTAQ TIVDNSSASADLV 590

Qy   540 VEDHTSNAQFVRMPDPNPFSSPTQPIIVNWTDLEVELHWNAPSTSGAGPITGYIIQYYS 599
       ||| | :| :| |:| |:| :| :| :| :| :| :| :| :|
Db   591 -----VRGP-----PGPPGGIRIEEIRDTAVALTWGR-GTDNHSPISKYITQ-SK 633

Qy   600 PDLGQTFWNPIDYVASTE-----YRIKGLKPSHSMFVIRAENEKGIGTPSVSSA 649
       |:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| :|
Db   634 TFLSEEWKD----AKTEPSDIEGNMESARVIDLIPWMYEFRIIATNTLG TGEPSPMSQ 688

Qy   650 LVYTSKPAAOVALSD-----KNKMDMAIAEK-----RL 677

```

Db 689 RIRTEGAPPNVAPSDVGGGGSNRELTITWMPLSREYHYGNNFGYIVAFKPFGEKEWRRV 748
Qy 678 T-----SEQLIKLE-----EV 688
Db 749 TVTNPEIGRYVHKDESMPPSTQYQVKVAFNSKGDGPFSLTAVIYSAQDAPTEVPTDVS 808
Qy 689 KTINSTAVRLFVKRKLLELIDGYIYK-WRGEPRTNDNQYVNVTSPTSTENY--VVSNLMP 745
Db 809 KVLSSSEISVSW-HHVTEKSVEGYQIRYWAADKEAAAQRVQV---SNQEYSTKLENLKP 864
Qy 746 FTNYEFFVIPYHSGVHSIHGAPSNMVDLTAAAPPSLPPEDVIRMLNLTTLR-----I 799
Db 865 NTRYHIDVSAFNS---AGYGPSPRTIDIITRAAPPSQRP---RI---ISSVRSGSRYII 914
Qy 800 SWKAPKADGINGILKGFQIVIVGQAPNNRNLTITNERAASVTLFHLVTGMTYKIRVAARS 859
Db 915 TWDHVKAMSNEHAVEGYKVLYRPDQHEGKLESTGKHTIEVP---VPSDEYVVEVRAHN 971
Qy 860 NGGVG 864
Db 972 EGGDG 976

Search completed: January 22, 2001, 12:26:10
Job time: 2047 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:59 ; Search time 162.41 Seconds
(without alignments)
257.899 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYLGFYHTHTHTYINFD.....TAQRFRSIPRNNIGVTQEQT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	674	9.8	1493	1	NEO1_MOUSE P97798 mus musculus
2	642	9.4	1377	1	NEO1_RAT P97603 rattus norv
3	639.5	9.3	1461	1	NEO1_HUMAN Q92859 homo sapien
4	632.5	9.2	1443	1	NEO1_CHICK Q90610 gallus gall
5	598	8.7	2012	1	DSCA_HUMAN O60469 homo sapien
6	597	8.7	1010	1	CONT_CHICK P14781 gallus gall
7	595	8.7	1020	1	CONT_MOUSE P12960 mus musculus
8	593	8.6	1018	1	CONT_HUMAN Q12860 homo sapien
9	585	8.5	1040	1	AXO1_RAT P22063 rattus norv
10	583.5	8.5	1257	1	CAML_HUMAN P32004 homo sapien
11	582.5	8.5	1040	1	AXO1_HUMAN Q02246 homo sapien
12	567.5	8.3	1260	1	CAML_MOUSE P11627 mus musculus
13	562.5	8.2	1036	1	AXO1_CHICK P28685 gallus gall
14	561	8.2	1284	1	NRCA_CHICK P35331 gallus gall
15	557	8.1	1259	1	CAML_RAT Q05695 rattus norv
16	554	8.1	1239	1	NRG_DROME P20241 drosophila
17	554	8.1	1447	1	DCC_HUMAN P43146 homo sapien
18	552	8.0	1447	1	DCC_MOUSE P70211 mus musculus
19	526	7.7	1897	1	PTPF_HUMAN P10586 homo sapien
20	509.5	7.4	1912	1	PTPD_HUMAN P23468 homo sapien
21	467	6.8	2029	1	LAR_DROME P16621 drosophila
22	462.5	6.7	1092	1	NCA2_XENLA P36335 xenopus lae
23	459.5	6.7	1088	1	NCA1_XENLA P16170 xenopus lae
24	455.5	6.6	1091	1	NCA1_CHICK P13590 gallus gall
25	455	6.6	1266	1	NGCA_CHICK Q03696 gallus gall
26	441.5	6.4	1070	1	PTK7_HUMAN Q13308 homo sapien
27	435.5	6.3	1115	1	NCA1_MOUSE P13595 mus musculus
28	433.5	6.3	853	1	NCA1_BOVIN P31836 bos taurus
29	433	6.3	858	1	NCA1_RAT P13596 rattus norv
30	432	6.3	848	1	NCA1_HUMAN P13591 homo sapien
31	432	6.3	1051	1	PTK7_CHICK Q91048 gallus gall
32	429	6.3	761	1	NCA2_HUMAN P13592 homo sapien
33	427.5	6.2	725	1	NCA2_MOUSE P13594 mus musculus

34	425.5	6.2	3707	1	PGBM_MOUSE Q05793 mus musculus
35	423.5	6.2	837	1	NCM2_HUMAN O15394 homo sapien
36	422	6.2	811	1	FS22_DROME P34083 drosophila
37	422	6.2	4393	1	PGBM_HUMAN P98160 homo sapien
38	409.5	6.0	837	1	NCM2_MOUSE O35136 mus musculus
39	407	5.9	873	1	FS21_DROME P34082 drosophila
40	398	5.8	1271	1	MYPC_CHICK Q90688 gallus gall
41	396	5.8	1913	1	KMLS_HUMAN Q15746 homo sapien
42	393	5.7	898	1	FAS2_SCHAM P22648 schistocerc
43	392.5	5.7	2481	1	UN52_CAEEL Q06561 caenorhabdi
44	381.5	5.6	1906	1	RMLS_CHICK P11799 gallus gall
45	370	5.4	1131	1	MYPF_CHICK P16419 gallus gall

ALIGNMENTS

RESULT 1
NEO1_MOUSE
ID NEO1_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEO1 OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE-BRAIN;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis."
RL Oncogene 15:691-700(1997).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UBIQUITOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y09535; CAA70727.1; -.
DR HSP; P02751; 1TTG.
DR MSD; MGI:1097159; NEO1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFM; PF00041; fn3; 6.
DR PFM; PF00047; ig; 4.

```

Qy 497 -----DSRISQHSQTGLSLHIAADLKKPDTGVYTCTIACNEDGE 531
      | : | | | : | | | : | | : |
Db 592 TPLSGNGEIQNYKLYMEKGTDKQDIDVSSH---SYTINGLKKYTEYSFRVVAYNKHGP 643

Qy 532 STWSASLTVEDHTSNAQFVRMPDPSPNFPSSPTQIPIINVNTDTE--VELHWNAP--STSGAGP 589
      : | : | : | : | : | : | : | : |
Db 649 GV-----STQDQAVARTLSDVPSAAPQNLSEVRNSKSIIVHWQPPSSTTQNGQ 696

Qy 590 ITGYIIQY----- 597
      ||| | : |
Db 697 ITGKYKIRYKASRKSVDTEITLVGTGQLSQLIEGLDRGTEYNFRVAALTVNGTGPATDWLS 756

Qy 598 ---YSPDLQQTWFNPIDYVASTE----- 617
      : | : | : | : |
Db 757 AETFESDLDE?--RVPEVPSSLHVRPLVTSIVVSWTPPENQNIIVVRGYAIGYGIGSPHAQ 814

Qy 618 -----YRIKGLKPSHSMYFIRAENEKGIGTSPVSVALVTTSKPAAQVALSDKNK 667
      | : | : | : : | | | | | : : | : | : |
Db 815 TIKVDYKQRYTYIENLDPSSHYVITLKAFFNVGEGIPLYESAV---TRPH-----TDTSE 866

Qy 668 MDMAIAEKRL?-----SEQLIK--EEVKTIINSTAVRLFWRKKRL---EELIDG--YYIKW 716
      : | : : : : : : : : : | : | : | : |
Db 867 VDLFVINAPYTPVPDPTPMMPVGVQASILSHDTRITWADNSLPKHQKITDSRYTIVRW 926

Qy 717 RGPRTN---DNQYVNVTPSPSTENYVVSNLMPFTNYEFFVI---PYHSGVHSI--HGAPS 768
      : | : : | : : | : | : | | | : | : | | |
Db 927 ---KTNPANTIKYKNANA--TTLSYLVGLKPNLTLYEFSVMVTKGRSSSTWSMTAEGA-- 979

Qy 769 NSMDVLTAEAPPSLPPEVDRI---RMLNLTLTRISWKAPKADGINGILKGFQIVIVQAPN 826
      : | : | : | : | : : : | : : : | : | : |
Db 980 -----TFELVPTSPPKDVTIVSKEGKPRTIIVNQPPSE--ANGKITGY--IIYSTDVN 1030

Qy 827 NNRNITTNERAASVTLFHLVTGMT---YKIRVAARSNGGVGVSHTSEVIMNQDTLEKH 882
      : | : : | : : : | : | : | : | : | : | : |
Db 1031 AEIHWDVIEPVVGNRLTHQIQELFLDTPPYFKIQRANSKGMG---PMSEAVQFR--TPKAD 1086

Qy 883 LAAQENESFLYGLINKSHVP----- 911
      : | : | : | : | : | : | : | : | : |
Db 1087 SSDKMPNDQALGSAGKSRPLDGLSKPPMSGGSNSPHGSPTSPLDSNMLLVIIVSVGVI 1146

Qy 912 IIFVVIIIAICYWRNSRNSDGRSFIKINDGSVHMASN-----NLW-----DV 955
      | : : : | : : : | : : : | : | : | : | : |
Db 1147 TIVVVVIAV?CTRRTTSHQKKRAACKSVNGSHYKGNCKDVKPPDLWIHHERLELKPI 1206

Qy 956 AQNPQNQNPMYNTAGRMTMNNRNGQALYSLTP--NAQDFNNCCDYSGTMHRPGSEHHYHY 1013
      : | : | : | : | | | : | : | : | : | : |
Db 1207 DKSPDPNPVMTD---TPIPRNSQ---DITPVDNSMD-----SNIHQRRNSYRGHE 1250

Qy 1014 AQLTGGPGNAMSTFYGNQYHDDPSPYATTTVLVSNQQPAWLNDKMLRAPAMPTNPVPPPEP 1073
      : : : : | : | : : | : | : | : | : | : |
Db 1251 SE-----DSMSTLAGRR-----GMRPKMM---MPFDSQPPQP 1279

Qy 1074 ----PARTADHTAGRRSRSSRASDGRGLT-----NGGLHHRTSGSQRSDSPPH 1117
      | : : : | : | | | : | : | : | : | : |
Db 1280 VISAHPHISLNDPNHHFHSSSLASPARSHLYHPSSWPIGTSMLSDRANSTESVRNTPS 1339

Qy 1118 TDVSYVQ-----LHSSDGTGSSKERTGERRTP-----PNKTLMDFIPPPPS 1158
      || : : : | : | : | : | : | : | : |
Db 1340 TDTMPASSSQCTCDHQDEPGATSSSYLASSQEEDSGQSLPTAHVRPSHPLKSFAPVA-- 1397

Qy 1159 NPPPPGGHYD?TATRRLNLRGSTP-----REDTYSVSDGAFARVDYNARPTSRNRNL 1211
      ||| | : | | | : | : | : | : | : |
Db 1398 -IPPPGPLYDPAL-----PSTPLLSQALEPSTFHSVKTASIGTLG--RSRPP----- 1443

Qy 1212 GGRPLGKRDDDSQRSSLLMDDDGGSSEAD 1241
      | : : : | : | | |
Db 1444 --MPVVVP?SAFEVQETTRMLEDESSSEYED 1471

```

RESULT 2
NEO1_RAT
ID NEO1_RAT STANDARD; PRT; 1377 AA

AC P97603;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR (FRAGMENT).
GN NE01 OR NGN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97015074; PubMed=8861902;
RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-Y.,
RA Culotti J.G., Tessier-Lavigne M.;
RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
RL Cell 87:175-185(1996).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U68726; AAB41100.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal.
FT NON_TER 1 1
FT SIGNAL <1 2 POTENTIAL.
FT CHAIN 3 1377 NEOGENIN.
FT DOMAIN 3 1074 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1075 1095 POTENTIAL.
FT DOMAIN 1096 1377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 36 105 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 135 197 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 232 296 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 324 386 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 405 502 FIBRONECTIN TYPE-III.
FT DOMAIN 505 598 FIBRONECTIN TYPE-III.
FT DOMAIN 599 698 FIBRONECTIN TYPE-III.
FT DOMAIN 704 798 FIBRONECTIN TYPE-III.
FT DOMAIN 819 919 FIBRONECTIN TYPE-III.
FT DOMAIN 920 1021 FIBRONECTIN TYPE-III.
FT DOMAIN 1087 1090 POLY-VAL.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 878 878 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1377 AA; 150637 MW; E514ED8ADD1A63A9 CRC64;

Query Match 9.4%; Score 642; DB 1; Length 1377;
Best Local Similarity 22.0%; Pred. No. 9.4e-28;
Matches 325; Conservative 194; Mismatches 535; Indels 426; Gaps 64;

Qy 36 PIDVVVSRGSPATLNCGA--KPSTAKITWYKDGQPVITNKEQV-NSHRIVLDTGSLFLK 92
Db 28 PVDTLVSRGSSVILNCAYSSEPS-PNIEWKKDG---TFLNLVSDRRQLLPDGLFISN 82
Qy 93 VNSGKNGKDSAGAYCYVASNEH-GEVKSNEGSLKLAHLREDFRVRPTVQALGGEMAVL 151
Db 83 VVHSKHNK-PDEGFIQCVATVDNLGTIVSRTAKLAVAGLPR-FTSQPESSIIYVGNISGL 140
Qy 152 ECSPPRGFPPEPVVSWRDKDELRIQDMPTYLHSDGNLIIDPVDSDSGTYQCVANNMVG 211
Db 141 NCEVNADL-VPFVRWEQNRQLLLDD--RIVKLPSGTLVISNATEGDEGLRCIVESGGP 197
Qy 212 ERVSNPARLSVFEPKPK-----FEQEPKDMTVDVGAALFDCRVTGDPQPIQITWKRKNP 265
Db 198 PKFSDEAELKVLQSEEMLDLVFLMRPSSMIKVIQSAVLPCVASGLPAPVIRW-MKNED 256
Qy 266 MPVTR-----AYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPPSFQT 320
Db 257 VLDTESSGRLALLA--GGSLSDVTEDDAGTYFCVADGNKTIIEAQELTVQVPEFLK 314
Qy 321 KPADQSPVAGGTATFECTLVGQSPAYFWSKEGQDQLLPSSYVSADGRTKVSPGTGLTIE 380
Db 315 QPANIYARESMDIVFECEVTGKRPATVWKVNG--DVVIPS-----DYFKIYKEHNLQVL 367
Qy 381 EVRVQDEGAYVCAGMSAGSSLSKA---ALKATFETKGRVKRKSMMGKQKQKVNQSIK 437
Db 368 GLVKSDEGFIQCAENDVGNQAQGLIIEHAPATTGPLSAPRDV-----VASLVS 420
Qy 438 YLISAVTGNTPAKPP-----PTIEGHQVN---QTLMVGSAIL 472
Db 421 TRFKLTWRTPASDPHGDNLTVSYFYTKEGVARERVENTSQPGEMQVTIQNLMPATVYIF 480
Qy 473 PCQASGK-----PTPGI-SWLRDGLPIDIT----- 496
Db 481 KVMQNKHGSGESSAPLRVETQPEVQLGCPAPNIRAYATSPTSITVTWETPLSGNGEIQN 540
Qy 497 -----DSRISQHSQSGSLHIADKKPDTGVYTCIAKNEDGESTVSASLTVED 542
Db 541 YKLYMEKGTQKEQDVVSSH---SYTINGLKTYEYFRVYVAINKHGPGV----- 558
Qy 543 HTSNAQFVRMPDPSNFPSSQTPPIVNVDTDE-VELHWNAPSTGA-GPITGYIIQY--- 597
Db 589 ---STQDVAVRTLSDVPSAQPQLSLEVRNKSIVIHWPQSSATQNGQITQYKIRYKA 645
Qy 598 -----YSPDLGQT 605
Db 646 SRKSDVTETVVTGQLSQLIEGLDRGTEYNFRVAALTVNGTGPAWDLSAETFESDLDES 705
Qy 606 WFNIPDYASTE-----Y 618
Db 706 --RVPEVPSLHVRPLTVSIVVSWTPPENQNIIVRGYAGYIGIGSPHAQTIKVDYKQRY 763
Qy 619 RIKGLKPSHSYMFVIRAENEKGIGTPSVSSALVTTSKPAAQVALSDKNKMDMAIEKRLT 678
Db 764 TIENLDPSSHYITLKAFNNVGEIPLYESAV---TRPH-----TDTSEVDLFVINAPYT 815
Qy 679 -----SEQLIKL-EEVKTINSTAVRLFWKRLK---EELIDG---YIKWRGPPTN---D 724
Db 816 PVPDPTPMPVPVGVQASILSHDTIRITWADNSLPKHQKITSRYTVRW---KTNIPIAN 871
Qy 725 NQYVNVTPSPSIENYVSNLMPTNYEFFVY---PYHSGVHSI--HGAPSNMMDVLTAEAP 779
Db 872 TRYKNANA-TLSYLVTLKPNLTLYEFSVMVTKGRSSSTWSMTAHA-----TFELV 922
Qy 780 PSLPPEDVRI--RMLNLTLRISWKAPKADGINGILKGFQIVIVGQAPNNNRNITTNERA 837
Db 923 PTSPPKDVTIVVSKGKPRTIIVNQPPSE--ANGKITGY-IIYYSTDVNAEIHWDVIEPV 979
Qy 838 ASVTILFHLVTGMT---YKIRVAARSNGGVG-VSHGTSEVIMNQDTLEKHLAAQENESF 892
Db 980 VGNRLTHQIQZLTLDTPYFFKIQARNKSGMGPMSEAVQFRTPKADSSDKMPNDQALGSAG 1039

Qy 893 LYGLI-----NKSH-----VPVIVIVAILIIFVVIITAYCWRN 926
Db 1040 KGRRLPLDGLSDYKPPMGSNSPHGSPTPLDNLMLLVIIISIGITVIVVVIITAVFCRR 1099
Qy 927 SRNSDGKDRSFIKINDGSVHMASN-----NLW-----DVAQNPNQNPMTYAGR 970
Db 1100 TTSHQKKRAACKSVNGSHKYGKNCCKVKKPDLWIHHERLELKPIDKSPDPNPMVMTD--- 1156
Qy 971 MTMNNRNGQALYSLTP--NAQDFNNDCCDYSMTMRPGSEHHYHAQLTGGPGNAMSTFY 1028
Db 1157 TPIPRNSQ---DITPVDNSMD-----SNIHQRRNSYRGHESE-----DSMSTLA 1197
Qy 1029 GNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPEPPARYADHTAGRSRS 1088
Db 1198 GRR-----GMRPKM-----MPFDSQPPQSVRNPSTDTMPASS 1232
Qy 1089 SRASDGRGTNGLLHRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTGERRTPPNKT 1148
Db 1233 S-----QTCTDHQDPEGATSSSYLASSQEDSGQLPTAHVR---PSHP 1274
Qy 1149 LMDFIPIPPSPNPPPGGHVYDTA-----TRQLNRGSTPREDTYSVSDGAFARVDVN 1201
Db 1275 LKSAFVPA---IPPPGPIYDPAIPSTPLLSQALNH---HLHSVKTASICTLGR--- 1323
Qy 1202 ARPTSRNRNLGGRPLKGRDDDSQRSSLMDDGGSSEAD 1241
Db 1324 SRPP-----MPVVVPSAPEVQEAATRMLEDSESSYEED 1355

RESULT 3

NEO1_HUMAN

ID NEO1_HUMAN STANDARD; PRT; 1461 AA.
AC Q92859; O00340;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEO1 OR NGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL; BRAIN;
RX MEDLINE=97236653; PubMed=9121761;
RA Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT "Identification and characterization of neogenin, a DCC-related
RT gene.";
RL Oncogene 14:1129-1136(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=FETAL BRAIN;
RX MEDLINE=97312699; PubMed=9169140;
RA Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA Korenberg J.R., Dreyer W.J.;
RT "Molecular characterization of human neogenin, a DCC-related protein,
RT and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT q23.";
RL Genomics 41:414-421(1997).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC LINES.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U61262; AAB17263.1; -.
DR EMBL; U72391; AAC51287.1; -.
DR MIM; 601907; -.
DR HSSP; P02751; 1TTG.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1461 NEOGENIN.
FT DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1106 1126 POTENTIAL.
FT DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 67 136 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 166 228 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 263 327 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 355 417 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 436 533 FIBRONECTIN TYPE-III.
FT DOMAIN 536 629 FIBRONECTIN TYPE-III.
FT DOMAIN 630 729 FIBRONECTIN TYPE-III.
FT DOMAIN 735 829 FIBRONECTIN TYPE-III.
FT DOMAIN 850 950 FIBRONECTIN TYPE-III.
FT DOMAIN 951 1052 FIBRONECTIN TYPE-III.
FT DOMAIN 1118 1121 POLY-VAL.
FT DISULFID 74 129 BY SIMILARITY.
FT DISULFID 173 221 BY SIMILARITY.
FT DISULFID 270 320 BY SIMILARITY.
FT DISULFID 362 410 BY SIMILARITY.
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 1248 1300 MISSING (IN ISOFORM 2).
FT CONFLICT 168 168 G -> N (IN REF. 2).
SQ SEQUENCE 1461 AA; 159958 MW; 7AAE897E69635A21 CRC64;

Query Match 9.3%; Score 639.5; DB 1; Length 1461;
Best Local Similarity 21.6%; Pred. No. 1.4e-27;
Matches 327; Conservative 202; Mismatches 546; Indels 437; Gaps 65;

Qy 36 PIDVVSRRGSPATLNCGA--KPSTAKITWYKDGQPVITNKEQV-NSHRIVLDTGSLFLK 92
Db 59 PVDTLVSRRGSSVILNCAYSSEPS-PKIEWKDG----TFLNLVSDRRQLPLDGLSIFSN 113
Qy 93 VNSGKNGKDSAGAYCYVSAEHE-GEVKSNEGSLKMLREDFRVRPRTVQALGGEMAVL 151
Db 114 VVHSKHNK-PDEGYQCVAIVESLGTIISRTAKLVAGLVR-FTSQPESSVYAGNGAIL 171
Qy 152 ECSPPRGFPPEVVSRRKDDKELRIQDMPTLHSDGNLIIDPVRSDSGTYQCVANNMVG 211
Db 172 NCEVNADL-VFVRWEQNRQPLLLDD--RVIKLPSGMLVSNATEDGGGLRCVVSSEGGP 228
Qy 212 ERVSNPARLSVFEKPK-----FEQPKDMTVDVGAALVDFCRVTGDPQPOQITWKRKNP 265
Db 229 PKYSDEVELKVLDPDEVISDLVFLKQPSPLVRVIGQDVLPCVASGLPTPTIKWKNKEA 288
Qy 266 MPV--TRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPPSFQTKPA 323

Db 289 LOTESERLVLLAGGSLEISDVTEDDAGTYFCIADNGNETIQAQLVQAQPEFLKQPT 348
Qy 324 DQSVPAAGTAIFECTLVGQSPAYFWSKEGQDLFPSTVSADGRTKVSPTGLTIEEVR 383
Db 349 NIYAHESMDIVFECEVTGKPTPTVKVKNKNG--DMVIPS-----DYPKIVKEHNLQVLGLV 401
Qy 384 QVDEGAYVCAGMNSAGSSLSKAL-----KATFET 413
Db 402 KSDEGFYQCIENDVGNAQAGAQLIILEHAPATTGPLSPAPRDVVASLVSTRFKLTWRT 461
Qy 414 KG-----RVQKKSKMKQKQKN-----VQSIKYKLISAVT-----GN 446
Db 462 PASDPHGDNLTYSVFTYKGIARERVENTSHPGEMQVTIQNLMPATVYIFRMAQNKHGS 521
Qy 447 TPAKPPPTIEHGHQNTLMVGSAILPCQASGKPTPGISW-----LRD 489
Db 522 GESSAPLRVE--TQPEVQLPGPAPNLRAYASPTSIIVTWETPVSGNGEIQNKLYIMEK 579
Qy 490 GLPIDITDSRISQHSSTGLHIADLKPKPTGVYTCIAKNEDGESTVASLTVEDHTSNAQF 549
Db 580 GTDKE-QQVDVSSH---SYTINGLKXYEYSFRVAVNKHGPGVSTPDVAVR-----627
Qy 550 VRMPDPSNFPSSPTQPIVNVTDTE-VELHWN--APSTSGAGPITGYIIQY-----597
Db 628 ---TLDVPSAAPQNLSEVRNKSIMIHQWPPAPATON-QGITGYKIRYKASRKSQDV 682
Qy 598 -----YSPDLGQTFWNIPD 611
Db 683 TETLVSGTQSLIEGLDRGTEYNFRVAALTINGTGPATDWSAETTESDLEDT--RVPE 740
Qy 612 YVASTE-----YRIKGLK 624
Db 741 VPSSLHVRPLVTSIVVSWTPPENQINIVRGYAIYGIGSPHAQTIKVDYKQRYTIENTLD 800
Qy 625 PSHSYMFIARAENEGIGTPSVSSALVTSKPAQVALSDKNKMDMAJAEKRLT-----S 679
Db 801 PSSHYITLKAFFNVGEGIPLYESAV---TRPH-----TDTSEVDFVINAPYTPVPDPT 852
Qy 680 EQLIKL-BEVKTINSTAVRLFVKRRKL---EELIDG--YIKWKGPPPTN---DNQYVNV 730
Db 853 PMMPVGVQASILSHDTIRITWADNSLPKQKITDSRYTYVW---KUNIPANTYKYNA 908
Qy 731 TSPSTENTYVSNLMPFTNYEFFVI---PYHSGVHSI--HGAPSNMMDVLTAEAPPSPPE 785
Db 909 NA-TTISLYTLGLKPNLTYEFSVMVTKGRSSSTWSMTAHTG-----TFELVPTSPPK 959
Qy 786 DVRI--RMLNLTLRISWKAPKADGINGILKGFQIVIVQAPNNRNITTNERAASVTLF 843
Db 960 DVTVVSKEGKPKTIIVNWQPPSE--ANGKITGY-IYYSTVDNAEHDWIEPVPVGNRLT 1016
Qy 844 HLTGMT-----YKIRVAARSNGGVSGHGTSEVIM-----NQDTLEKHLAAQENESF 892
Db 1017 HQIQELTLDTPYFYKIQARNKGMG---PMSEAVQFRTPKADSSDKMPNDQASGSGGKGS 1073
Qy 893 LYGLINKSHVP-----VIVIVAILIIFVVIITAYCYWRNSRN 929
Db 1074 RLPDLGSDYKPPMSGNSPHGSPSPLDSNMLLVIVSVGVITVIVVIAVFCRTRTTS 1133
Qy 930 SDGKDRSFIKINDGSVHMASN-----NLW-----DVAQNPNQPNMYTAGRTM 973
Db 1134 HQKKRAACKSVNGSHYKGNKSKDVKPPDLWIHHERLEKPKIDKSPDPNPIMTD---TP 1189
Qy 974 NNRNGQALYSLTP--NAQDFPNCCDYSGMTHPGSEHYHYAQLTGSGPNAMSTFYGNQ 1031
Db 1190 IPRNSQ--DITPVDSMD-----SHIHQRNRYRGHESE-----DSMSTLAGRR 1231
Qy 1032 YHDDPSYATTLVLVSNQPAWLNDKMLRAPAMPTNPVPEP-----PARYADHTAGRRS 1086
Db 1232 -----GMRPKMM---MPFDSQPPQVISAHPHSLDNPHHFH 1266
Qy 1087 RSSRASDGRGTL-----NGGLHRTSGSQRSDSPPHTDVSYVQ-----1124

Db 1267 SSSLASPARSHLYHPGSPWPIGTSMSLSDRANSTESVRNTPSTDTMPASSSQCTCDHQD 1326
Qy 1125 ---LHSSDGTGSSKERTGERRTP-----PNKTLMDFIPIPPPPPPGGHYDTA-----1171
Db 1327 PEGATSSSYLASSQEDSGQLPTAHVRPSHPLKSFAPVA---IPPPGPPYDPALPSTP 1383
Qy 1172 --TRQLNRGSPREDTYDSVSDGAFARVDVNARPTSRNRNLGGRPLKGRDDDSQRSSL 1229
Db 1384 LLSQALNH---HIHSVKTASIGTLGR---SRPP-----MPVVPSAPEVQETTR 1427
Qy 1230 MMDDDGGSSEAD 1241
Db 1428 MLEDSSESYEPD 1439

RESULT 4
NEOL_CHICK
ID NEOL_CHICK STANDARD; PRT; 1443 AA.
AC Q90610;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE-95105243; PubMed-7806578;
RA Vielmetter J., Roman J.M., Dreyer W.J.;
RT "Neogenin, an avian cell surface protein expressed during terminal
RT neuronal differentiation, is closely related to the human tumor
RT suppressor molecule deleted in colorectal cancer.";
RL J. Cell Biol. 127:2009-2020(1994).
CC -|- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -|- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -|- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
CC AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC -|- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -|- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07644; AAC59662.1; -.
DR HSSP; P80362; !NTL.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; lg; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN <1 1090 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1091 1111 POTENTIAL.
FT DOMAIN 1112 1443 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 102 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 132 194 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 229 293 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 321 383 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 422 519 FIBRONECTIN TYPE-III.
FT DOMAIN 522 615 FIBRONECTIN TYPE-III.


```

Qy 36 PIDVWVSRGSPATLNCAGKPS- AKITWYKDGQPVITNKEQVNSHRIVLDTGSLFLKLVN 94
   |::: ||: |: | | | | | : | : | | | : |
Db 25 PMDILSVRGASVIMNCSSYCETPPKIEWKDG--TLNLVLS--DDRQLLPDGSLLINSV 81

Qy 95 SGKNKGDSAGAYYCVASNEH-GEVKSNEGSLKAMLREDFVRVPRTVQALGGEMAVLEC 153
   |: | | | | | | | | : | | | | | : | | | |
Db 82 HSKHNK-PDEGYQCVAIVESLGSIVSRAKLTVAGLPR-FTSQPELSSVYKGSAILNC 139

Qy 154 SPPRGFPPEPVWSWRKDDKELRIQDMPRTYTLHSDGNLIIDPVDRSDSGTYQCVANNMVG 213
   | | | | : | : | | | | | | | : | | | | : | :
Db 140 EVNVDL-APFVRWEQDRQPLSLDD--RVFKLPSPGALLIGNATDTDGGFYRCVIESGGTPK 196

Qy 214 VSNPARLSVFEKPK-----FEQEPKMDTVDGAAVLDFCDRTGDPQPOPTWKRKNEMP 267
   | | | : | : | | : | | | | | | | | | | : |
Db 197 YSEEAELKILPDPEEPQSLVFEVRQPSLLTKVTGQNAVFPVAGGFPPTPYRWTKNGEELI 256

Qy 268 V--TRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPSPFQTKPADQ 325
   : : : | | | | | | | | | | | | | | | | |
Db 257 TEDSERFALRAGSLLISDVTEEDVGITYTCIADNENETIEAQELAVQVPPEFLKRPANI 316

Qy 326 SVPAGGTATFECTLVGQSPAYFWSKEGQDLLFSPYSVADGRTKVSPGTGLTIEVRQV 385
   | | | : | | | | | | | | | | | | | | : | :
Db 317 YAHESMDIVFEEVTKPTKPTVKWKNG--DVVIPS-----DYFKIVKEENLQVLGLVKS 369

Qy 386 DEGAYVCAGMNSAGSSLSKAAL-----KAT----- 410
   | | | | | : : : | | | | | | | |
Db 370 DEGFYQCIAENDVNGAQAQAGLIILDLVAIPTLPPTSLSATNDHLAPATTGPLTAPR 429

Qy 411 -----FETKGRVQKKK-----SKMKQKQKNVQS 434
   | | | | : : : | | | | | | | |
Db 430 DVVATLVSTRFIRLTWRTVPSPDQGNLTYSIPTYKEGINRERVENTSRG-ETQVNIQN 488

Qy 435 IIK--YLISAVTGNTPAKPPPTIEHGHQNTLMWGSSAILPCQASGKPTGISWLRDL 491
   : : | : | | : | | : : : | : | | | |
Db 489 LMPETVYVFRVVAQN-----KHGHGESSAPLKVATQPEVQLPG-PAPNIR-AYAGS 537

Qy 492 PIDIT---DSRIS-----QHSTGSLHIADLKFPDVG-----YTCLIA 525
   | : | : : | | : | | : | | : | | : | :
Db 538 PTVSYVTWETPLSGNGEIQNYKLYMEKGQDSEQDQVDVAGLSYITGLKKYTEYSFRVVA 597

Qy 526 KNEDGESTWSASLTVEDHTSNAQFYRMPDPNPNFSSPTQPIIWNVTDT-VELHWN-APS 583
   | : | : | | : | | : | | : | | : | | : |
Db 598 YNKHGPGV-----STQDVVVRTLSDVPSAAPQNLTEARNSKSLIMHWQPPPA 645

Qy 584 TSCAGPITGIIQY----- 597
   : : | | | | | |
Db 646 GTHSGQITGYKIRYKRSKSDVTESVGGTQLFQLEGLERGTEYNFRIAMTVNGTGA 705

Qy 598 -----YSPDLGQTFWNPIDYASTE----- 617
   : | | | | : | : |

```

```

Db      706 TDWSAETFESDLDES---RVPEVPSSSLHVRPLVTISIVSWTTPPENQNIWVRGYAIGYICG 763
Qy      618 -----YRIKGLKPSHSYMVFIRAENEKGIGITSPVSVALVTTSKPAQVAL 662
          | :| :| :| :| :| :| :| :| :| :|
Db      764 SPHAQTIKVDYQRYYTIENLDPSSHYVITLKAFNNVGEGIPLYESA---TRPH----- 815
Qy      663 SDKNKMDMAIAEKRLT----SEQLIKL-EEVKTINSTAVRLFWKRRKL--EELIDG-- 711
          || :|| :| :| :| :| :| :| :| :| :| :|
Db      816 SDTSEVDLFVINAPYTPVPDPSMPMPYPGVQASILSHDTIRITWADNSLPKNQKITDARY 875
Qy      712 YYIKWRGPRTN--DNQYVNVTSPSTENYVSNLMPTNYEFFVI---PYHSGVHSI--H 764
          |:| :| :| :| :| :| :| :| :| :| :| :|
Db      876 YTVRW---KTNIPANTKYKTANATLSYLVTGLKPNTLYEFSVMVTKGRSSTWSMTAH 931
Qy      765 GAPNSMMDVLFAEAPPSPLPEDVRI--RMLNLTLTRISWKAPKADGINGILKFQIIVG 822
          :| :| :| :| :| :| :| :| :| :| :| :|
Db      932 GT-----TFELVPTSPPKDVTVVSKEGKPRTIIWNMQPPE--ANGKITGY-IIYYS 980
Qy      823 QAPNNRNITTNERAASVTLFLHLVTGMT---YKIRVAARSNGGVG----- 864
          :| :| :| :| :| :| :| :| :| :| :| :|
Db      981 TDNWAEHDWIEPVVGNRLTHQIQLELTLDPTYYPFKIQARNSKGMGPMSAEVQFRTPKAE 1040
          :| :| :| :| :| :| :| :| :| :| :| :|
Qy      865 -----VSHGTSEVIMNQDTEKHLAAQENESFLYLINKSHVPVIVIVAIL 911
          :| :| :| :| :| :| :| :| :| :| :| :|
Db      1041 SSDKMPPDQASGSACKGSRPVDVGPDKPPLSGSNSPHGSPTSPLDSNMLLIVIIVSGVI 1100
Qy      912 IIFVVIIIAYCYWRNSRNSDGKDRSFIKINDGSVHMASN-----NLW-----DV 955
          | :| :| :| :| :| :| :| :| :| :| :| :|
Db      1101 TIVIVIVAVFTCRRTTSHQKKRAACKSVNGSHKYKGNCKDVKPPDLWIHHERLELKPI 1160
Qy      956 AQNPQNPNMYMITAGRMTNMNRNGQALYSLTP--NAQDFFNCCDYSGTMHRPGSEHHYHY 1013
          :| :| :| :| :| :| :| :| :| :| :| :|
Db      1161 DKSPDPNPIMTD---TPIPRNSQ---DITPVDNSMD-----SNIHQRRNSYRGHE 1204
Qy      1014 AQLTGPGNGAMSTFYGNQYHDDPSPYATTTLVLNSNOQPALNDKMLRAPAMPTNPVPPPEP 1073
          :: :|| :| :| :| :| :| :| :| :| :| :|
Db      1205 SE-----DSMSTLAGRR-----GMRPKMM--MPPDSQQPPQ 1233
Qy      1074 ----PARYADHTAGRRSSRSSRASDRGRTLNGGLHHRTSGSQRSDSPPHD----- 1119
          | :| :| :| :| :| :| :| :| :| :| :|
Db      1234 VISAHPIHSLDNPHHHFHSGSLASPSTRSY---LHHQVSPWPVGTSMShDRANSTESVR 1289
Qy      1120 -----VSYYQLHSSDGT--GSSKERTGER RTP-----PNKTLMDFIPPP 1156
          :| :| :| :| :| :| :| :| :| :| :| :|
Db      1290 NTPSSDTMPASSSQPCADHQPDSSSGAYLGSAQEEDAQSLETAHVVRPSHPLKSFAPV- 1348
Qy      1157 PSNPPPPGGGHYD-TATRRLNLRGSTPREDTYD--SVSDGAFAVRDVVNARPTRSNNRLGG 1213
          | :| :| :| :| :| :| :| :| :| :| :|
Db      1349 ---AVPAAGSAYDPLTLPSTPLLTQQAPSHPVHSVKTASIGTLGR---TRPP----- 1393
Qy      1214 RPLKGRKRDDDSQRSSLMMDDDDGGSSEAD 1241
          | :| :| :| :| :| :| :| :| :| :| :|
Db      1394 MPVVVPSADPVQOETTRLMEDESSSEYEP 1421

```

```

RESULT 5
DSCA_HUMAN
ID DSCA_HUMAN STANDARD; PRT; 2012 AA.
AC O60469; O60468;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE PRECURSOR (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendel M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;

```

FT	DOMAIN	885	972	FIBRONECTIN TYPE-III.
FT	DOMAIN	984	1076	FIBRONECTIN TYPE-III.
FT	DOMAIN	1088	1177	FIBRONECTIN TYPE-III.
FT	DOMAIN	1189	1273	FIBRONECTIN TYPE-III.
FT	DOMAIN	1300	1366	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	1380	1463	FIBRONECTIN TYPE-III.
FT	DOMAIN	1477	1562	FIBRONECTIN TYPE-III.
FT	DISULFID	46	102	BY SIMILARITY.
FT	DISULFID	145	197	BY SIMILARITY.
FT	DISULFID	246	293	BY SIMILARITY.
FT	DISULFID	335	385	BY SIMILARITY.
FT	DISULFID	428	484	BY SIMILARITY.
FT	DISULFID	525	575	BY SIMILARITY.
FT	DISULFID	617	669	BY SIMILARITY.
FT	DISULFID	711	766	BY SIMILARITY.
FT	DISULFID	809	865	BY SIMILARITY.
FT	DISULFID	1307	1359	BY SIMILARITY.
FT	CARBOHYD	28	28	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1562	1571	NFATLNYDGS -> KEAARCKEFS (IN SHORT ISOFORM).
FT	VARSPLIC	1572	2012	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	1893	2012	HRPGDLIHLHPPLYLRMDPELLNRGGPGTGRDLSLQACLEPQK
FT				SRTLKRPTVLEPIPMEEAASSASSTREGSQWQPGAVATLPQR
FT				GEAGELGQAARKSSSQSLDGRHLKGNPNYAKSYTLV ->
FT				IGQVTSYICLHLEWTFPC (IN REF. 1).
SQ	SEQUENCE	2012 AA; 222259 MW;	0E33CFB781A08334 CRC64;	

Query Match 8.7%; Score 598; DB 1; Length 2012;
Best Local Similarity 25.2%; Pred. No. 4.3e-25;
Matches 230; Conservative 131; Mismatches 365; Indels 188; Gaps 41;

```

Qy      16 YINFDPKIPNASNLAPVIEHPI-----DVVVSRSGPATLNCAGKAPS-TAKITWYKDG  66
      :: ||: |: : ::| : || : |: | : : || |
Db      386 FVRKDKL-SAQDYVQVVLDEGTPKIIISAFSEKVVSPAEPVSLMCKVGTPLTITWTLLD  444

Qy      67 QPVITNKEQVNSHRI---VLDTGSFL-LLKVNSGKNGKSDAGAYYCVASNEHEGVKSNE  122
      |: : ||| : |: : |: : |: | |: | |
Db      445 DPILKG---GSHRISQMITSEGNVSVYSLNISS---SQVRDGGVYRCTANNSAGV---  493

Qy      123 GSLKLAMLRDFR---VRP-RTVQALGGEMAVLECSPPRGFPEPVVSWRKDDKELRIQD  177
      | |: | |: | : |: | : |: | |: | |: |
Db      494 ---LYQARINVRGPASIRPMKNITAIAGRDYIHC-RVIGYPYISIKWYKNSNLPPFNH  548

Qy      178 MPRYTLHSDGNLIIDPVDRS-DSGTQYQCVANNMVGERSVN-----PARLSVFEKP  226
      : : |: | : : |: | : |: | : :| : |: |
Db      549 R-QVAFENNGILLKLSDVQKEVDEGEYTC--NVLVQPOLSTSQSVHVTVKVPFPIQPFEP  605

Qy      227 KFEQEPKDMTVDVGAAVLFD-C-RVTGDQPQITWKRKNEPMPTRAYIAKDN---RGLR  281
      |: | |: | |: | |: | |: | |: | |: |
Db      606 RF-----SIGQRVFIPCVVVSGDLPTITITWQDGRPIGSLG-VTIDNIDFTSSLR  655

Qy      282 IERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPPSFQTKPADQSVPAAGGTATFECTLVG  341
      : : |: | : : |: | : |: | : :| : |: |
Db      656 ISNLSLMHNGNYTCIARNEAAVAEHOSOLIVRVPKPFVQVPRDODGIGYKAVILNCSAEG  715

```

```
Qy      342 QSPAYFW--SKEGQDQLLFFSYVSADGRKTVSQPTGLTITIEVRQVDEGAYVCAGMNSG 399  
       | | | | | : : : | : | : | : | : |  
Db      716 YPVPTIWWFKSGKAGVPQFQP---IALNGRIQVLNSSLIKHVVEEDSGYILLCKVSNDBG 773  
  
Qy      400 SSLSKAALKATFETKGRVKQKKSKMKQKQKRNQSIKYILISAVTGNTPAKPPTTIEHGH 459  
       : || : | : | : | : | : | : | : |  
Db      774 ADVKS-----MYLT VKI-----PAMITSY 793  
  
Qy      460 QNOTLMV-GSSAILPCQASGKPTPGISWLRLGPLID-----ITDSRISQHTSGSLHIA 511  
       | | | | | : | : | : | : | : | : | : |  
Db      794 PNTTLATQGOKKEMSCTAHGKEPIIVRWEEKDRINPEMARYLVSTKEVGEEVISTLQIL 853  
  
Qy      512 DLKKPDTPGVYTICAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSPFPSSPTQPIIVNV 571  
       : : | : | | | | | | : | : | : | : |  
Db      854 PTVREDSGFFSCHAINSYGEDRGIIQLTVQE-----PPDPEEIEIKDVK 897  
  
Qy      572 DTEVELHWNAPSTSGAGPTGYIIQQYSPDLGQTWFNI-----PDYVASTEYRIKGLK 624  
       : | | : | : | | | | | : | : | : | : |  
Db      898 ARTITLRWTM-GPDGNPSITGYDIE--CRNKSDSWDSAQRKTVDVSQLNSAT---IIDIH 951  
  
Qy      625 PSHSYMFIARAENEKGIGTSPSVSSALVTTSKPAQVALSDKKNMDMAIAEKRLTSEOLIK 684  
       | | : | : | | | : | : | : | : | : |  
Db      952 PSSYSIRMYAKNR--IGKSEPSNELTITADEAAPDG-----PPQEVH 992  
  
Qy      685 LEEVKTIINSTAVRLFWK--KRKLEE-LIDGYIKWRGPPTNDNQYVN---TSPSTEN 737  
       | | : | : | : | : | : | : | : | : | : |  
Db      993 LE---PISSQSIRVTWKA PKKLHLONGIIRGYIGYR-EYSTGGNFQFNII SVDTSGDSEV 1048  
  
Qy      738 YVVSNLMPFTNYEFFVIPYHSGVHSIHGAPNSMDVLTA--EAPPSPPEPDVRI RMLNLT 795  
       | : | : | : | : | : | : | : | : | : |  
Db      1049 YTLDNLNKFTQYGLVVQCACNRA----GTGPSSQEIITTTLEDVPSYPPE NVQAIATSPE 1103  
  
Qy      796 TLRISWKAPKADGINGILKFQIVIVGQAPNNN---RNITTNERAASVTLFLHVTGMTY 851  
       : : | | : : | : | : | : | : | : | : |  
Db      1104 SISISWSTLSREALNGILQGFRIVYWANLMDGELGEIKNITTQ--PSLELDGLEKYTYN 1161  
  
Qy      852 KIRVAARSNGGVGV 865  
       | : | : | | |  
Db      1162 SIOVLAFTRAGEDV 1175
```

```

RESULT      6
CONT_CHICK
ID   CONT_CHICK      STANDARD;          PRT;  1010 AA.
AC   P14781; P10450;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CONTACTIN PRECURSOR (NEURAL CELL RECOGNITION MOLECULE F11).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
     Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90180453; PubMed=2627374;
RA   Bruemendorf T., Wolff J.M., Rainer F., Rathjen F.G.;
RT   "Neural cell recognition molecule F11: homology with fibronectin type
RT   III and immunoglobulin type C domains.";
RL   Neuron 2:1351-1361(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN;
RX   MEDLINE=89008597; PubMed=3049624;
RA   Ranscht B., Dours M.T.;
RT   "Sequence of contactin, a 130-kD glycoprotein concentrated in areas
RT   of interneuronal contact, defines a new member of the immunoglobulin
RT   supergene family in the nervous system.";
RL   J. Cell Biol. 107:1561-1573(1988).
RN   [3]
RP   GPI-ANCHOR.
RX   MEDLINE=89286606; PubMed=2735929;
RA   Wolff J.M., Bruemendorf T., Rathjen F.G.;

```

```

RT "Neural cell recognition molecule F11: membrane interaction by
CC covalently attached phosphatidylinositol.";
RL Biochem. Biophys. Res. Commun. 161:931-938(1989).
CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
CC SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE C-TERMINUS
CC AND IS LONGER DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14877; CAA33018.1; -.
DR EMBL; Y00813; CAA68753.1; ALT_FRAME.
DR PIR; JU0094; JU0094.
DR PIR; S01998; S01998.
DR HSSP; P20241; 1CFB.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 19
FT CHAIN 20 ? CONTACTIN.
FT PROPEP ? 1010 REMOVED IN MATURE FORM.
FT DOMAIN 50 113 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 143 210 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 247 308 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 336 389 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 420 482 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 510 581 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 593 599 GLY/PRO-RICH.
FT DOMAIN 600 701 FIBRONECTIN TYPE-III.
FT DOMAIN 702 803 FIBRONECTIN TYPE-III.
FT DOMAIN 804 900 FIBRONECTIN TYPE-III.
FT DOMAIN 901 996 FIBRONECTIN TYPE-III.
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1010 AA; 112507 MW; 2E38F071AF423AE1 CRC64;

```

Query Match 8.7%; Score 597; DB 1; Length 1010;
Best Local Similarity 23.6%; Pred. No. 1.8e-25;
Matches 268; Conservative 140; Mismatches 364; Indels 364; Gaps 47;

```

Qy      6 FYHTHTHTHTYI-----NFDKIPNASNLAPVIEHPIDVVV---SRGSPATLNCGAKP 55
      | : : | | | | : : | | | | : : | | :
Db      3 FFISHLVTLCFIFCVADSTHFSEEGN-KGYGPVFEEQPIDTIYPESSDQGVSMNCRAR- 60

Qy     56 STAKITWKDGPQVPTINKEQVNSHRIVL---DTGSLFLLKYNVSGKNGKSDAGAYYCVASN 113
      | | | : : | | : : | | | | |
Db     61 -----AVPFPTYKWLNNWDIDLTKDRYSVMGGRLLVISNPEKSRDACKYVCVSN 110

Qy    114 EHGEVKSNEGSLKLAML-----REDFRVRPRTVQALGGEMAVLECSPPRGFPEPV-VWSR 167
      | : | : | | | : : | | | | | : : |
Db    111 IFTGTVRSSEATLSFGTLDPPFPEEHYEVKVR---GVCAYLLCEPPYHPYDDLRYRL 165

Qy    168 KDDKELRIO-DMPYRTILHSDGNLIIDPVRDSGTYOCVANNMVGERSVNPA-RLSVFEK 225

```

Db 166 LNEFPFIALDRRRVSQTNGNLYIANVEASDKGNYSCF-----VSSPSITKSVFSK 217
Qy 226 ----PKFEQE---PKDMTVD-----VGAALVDCRVTGDPQQTITWRKNEPMPVT- 269
Db 218 FIPLIPQADRAKVYPADIKVKFKDTYALLQNVLTCECFALGNVPPELRWSKYLEPMPATA 277
Qy 270 -----RAYI-----AKD--- 276
Db 278 EISMSGAVLKIFNIQEDEGLYECEAENYKGDQKHQARVYVQASPEWVEHINDTEKDGS 337
Qy 277 -----NRG-LRIERVQPSDEGEYVCYARNPAGTLEASHL 310
Db 338 DLYWPCVATGKPIPTIRWLKNGVSPFRKGLRIQGLTFEDAGMYQCIENAGHIYANAE 397
Qy 311 RVQA-PPSFQTKPADQSVPA--GGTATFECTLVGQSPAYFWSKGGQDQLLPSPVVSADG 367
Db 398 KIVASPTTFELNPMKKKILAAKGGRIIECKPKAAPKPFWSK-GTELLVNGS----- 450
Qy 368 RTKVSPTGTLTIEVRQVDEGAYVACAGNSAGSSLSKAALKATFETKGRVQKKSKMGKQ 427
Db 451 RIHIWDDGSLIINVTKLDEGRYTCFAENNRKANSTGVLEMTATR----- 497
Qy 428 KQKNVQSIILYISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGIS-- 485
Db 498 -----ITLAPLVNDVTGGENATMQCIASHDPTLDTLPI 530
Qy 486 WLRDGLPIDIT-----DSRISQHSGLIADLKPDPTGYTICAKNEDGESTWSASLT 539
Db 531 WSLNGFVIDFEKEHEHYERNVMIKSNGELLIKQVLRHAGRYTCTAQTIVDNSSASADLV 590
Qy 540 VEDHTSNAQFVRMPDPSPNFPSSPTQPIIVNVTDTEVELHWNAPSTSGAGPTGYIIQYIS 599
Db 591 -----VRGP-----PGPPGGIRIEEIRDTAVALTWSR-GTDNHSPIKSYTIQ-SK 633
Qy 600 PDLQGTWFPNIDYVASTE-----YRIKGLKPSHSMYFVIRAEKNGIGTPSVSSA 649
Db 634 TFLSEWKD-----AKTEPSDIEGNMESARVIDLIPKMEYEFRIIATNTLTGTGEPSPMQ 688
Qy 650 LVTTSKPAQVALSD-----KNKMDMAIAEK-----RL 677
Db 689 RIRTEGAPPNVAPSDVGGGGGNSRELITITWPLSREYHYGNNGFYIVAFKPFGEKWRVR 748
Qy 678 T-----SEQLIKLE-----EV 688
Db 749 TVTNPEIGRYVHKDESMPSTQYQVVKAFNSKGDGPFSLTAVIYSAQDAPTEVPTDVS 808
Qy 689 KINSTAVRLFVKKKLEELIDGYYIK-WRGPRTNDNQYVNVTSPTENY--VVSNLM 745
Db 809 KVLSSSEISVSW-HVTEKSVEGYQIRYWAHDKAAARQVQV---SNQYESTKLENLAP 864
Qy 746 FTNYEFFVPIYHSGVHSIHGAPNSMDVLTAEAPSLPPEDVRIRMLNLTTLR-----I 799
Db 865 NTRYHIDVSAPNS---AGYGPSRTIDITRKAPPSQRP---RI-----ISSVRSGRYII 914
Qy 800 SWKAPKADGKINGILKGFQIVIVGQAPNNRNITTNERAASVTLFHLVTGMYKIRVAARS 859
Db 915 TWDHVKAMSNEASVEGYKVLYRPDQGHGKLFSTGKHTIEVP---VPSDEGYVVEVRAHS 971
Qy 860 NGGVG-----VSHGTSEVMNQDTLEKHLAAQENESPLYLINKSHVPIVIVA 909
Db 972 EGGDGEVAQIKISGATAGV-----PTLLGLV---LPALGVLA 1006

RESULT 7

CONT_MOUSE

ID CONT_MOUSE STANDARD; PRT: 1020 AA.

AC P12960;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CONTACTIN PRECURSOR (NEURAL CELL SURFACE PROTEIN F3).

GN CNTN1.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=BRAIN;
RX MEDLINE=89340657; PubMed=2474555;
RA Gennarini G., Cibelli G., Rougon G., Mattei M.-G., Goridis C.;
RT "The mouse neuronal cell surface protein F3: a phosphatidylinositol-
RT anchored member of the immunoglobulin superfamily related to chicken
RT contactin.";
RL J. Cell Biol. 109:775-788(1989).
CC -1- FUNCTION: MEDIATES CELL SURFACE INTERACTIONS DURING NERVOUS
CC SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- MISCELLANEOUS: F3 SHARES WITH L1, N-CAM, MAG, AND OTHER CELL
CC ADHESION MOLECULES FROM NERVOUS TISSUE THE L2/HNK-1 CARBOHYDRATE
CC EPITOPE.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14943; CAA33075.1; -.
DR PIR; S05944; S05944.
DR HSSP; P40189; 1BQU.
DR MGD; MGI:105980; CNTN1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 20
FT CHAIN 21 ? CONTACTIN.
FT PROPEP ? 1020 REMOVED IN MATURE FORM.
FT DOMAIN 58 121 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 151 218 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 256 317 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 345 398 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 429 491 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 519 592 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 604 611 GLY/PRO-RICH.
FT DOMAIN 611 712 FIBRONECTIN TYPE-III.
FT DOMAIN 713 814 FIBRONECTIN TYPE-III.
FT DOMAIN 815 910 FIBRONECTIN TYPE-III.
FT DOMAIN 911 1006 FIBRONECTIN TYPE-III.
FT DISULFID 65 114 BY SIMILARITY.
FT DISULFID 158 211 BY SIMILARITY.
FT DISULFID 263 310 BY SIMILARITY.
FT DISULFID 352 391 BY SIMILARITY.
FT DISULFID 436 484 BY SIMILARITY.
FT DISULFID 526 585 BY SIMILARITY.
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1020 AA; 113388 MW; 9DCDAA40EAA4CB7 CRC64;

Query Match 8.7%; Score 595; DB 1; Length 1020;
Best Local Similarity 24.0%; Pred. No. 2.3e-25;

Db 990 SOVKISGVSTLSSSL 1004

FT DOMAIN 813 908 FIBRONECTIN TYPE-III.

RT "The gene encoding L1, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [6]
RP SEQUENCE OF 1030-1257 FROM N.A.
RX MEDLINE=91132183; PubMed=1993895;
RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
RA Stallcup W.B.;
RT "Isolation and sequence of partial cDNA clones of human L1: homology
RT of human and rodent L1 in the cytoplasmic region.";
RL J. Neurochem. 56:797-804(1991).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE=88298876; PubMed=3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule L1.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP VARIANT HSAS TYR-264.
RX MEDLINE=94004956; PubMed=8401576;
RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnai D.;
RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANT HSAS/MASA LEU-1194.
RX MEDLINE=95187172; PubMed=7881431;
RA Fransen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the L1CAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE=95004608; PubMed=7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metzenberg A., Ionasescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the L1 gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE=95004609; PubMed=7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
RA Reyniers E., Korn B., Poustka A., Wilson G., Schrander-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT L1CAM.";
RL Nat. Genet. 7:408-413(1994).
RN [12]
RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
RX MEDLINE=95282776; PubMed=7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Wadelius C., Kenwick S.;
RT "New domains of neural cell-adhesion molecule L1 implicated in
RT X-linked hydrocephalus and MASA syndrome.";
RL Am. J. Hum. Genet. 56:1304-1314(1995).
RN [13]
RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
RX MEDLINE=96153146; PubMed=8556302;
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, L1.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]

RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in L1-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; V-439--T-443 DEL; C-784 AND L-936--L-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).
RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE=98112489; PubMed=9452110;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
RL Hum. Mutat. Suppl. 1:S284-S287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE=98415726; PubMed=9744477;
RA Saugier-Verber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
RT "Identification of novel L1CAM mutations using fluorescence-assisted
RT mismatch analysis.";
RL Hum. Mutat. 12:259-266(1998).
CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF THREE X-LINKED
CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
CC SYLVIVUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
CC Milder PRESENTATION AND A LONGER LIFE EXPECTANCY. 3: SPASTIC
CC PARAPLEGIA TYPE 1 (SPG1). COLLECTIVELY THESE SYNDROMES ARE ALSO
CC KNOWN AS CRASH SYNDROME, AN ACRONYM WHICH STANDS FOR CORPUS
CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,
CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF HIRSCHSPRUNG DISEASE
CC (HSCR).
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=L1CAM; NOTE=L1CAM mutation Web Page;
CC WWW="http://hgins.uia.ac.be/dnalab/11".
CC

[illegible]

```
Qy   684 -----KLEEVKTINSTAVRLFWKRRKLEEL---IDGYI 713  
      :||::|:||::|::|::|:  
Db   790 AVNSQGKGPEPQTIGYSGEDYPQAIPLEGIEILNSSAVLVKWRPVDLQVKGHLRGN 849  
  
Qy   714 IK-WR-GPPRINDNQYVN---VTSPSTENVVVSNLMPTTINYEFFVIPYHSGVHSIHGAP 767  
      : || | | :::: | :|::|::|::|::| | | :::|  
Db   850 VTYWREGSQRKHKSRKHDKHVVPANTTSVILSGLRPYS-----YHLEVQAFNGR 902  
  
Qy   768 SNSMDVLTAEPAPSLP--PEDVRIRMLNLTTLRIISWKAPKADGINGILKFQIVIVGQAP 825  
      | | | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db   903 SGPASEFTFSTPEGVPGHPALHLECQNSTLLLRWPPLSH--NGVLTYGVLSYHPLDE 960  
  
Qy   826 NNNRNITTNERAASV---TLFLHVTGMTYKIRVAARSNGGVG 864  
      : : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |  
Db   961 GGGQQLSFNLRDPELRTHNLTDLSPHLYRFQIQATTKEGPG 1002
```

RESULT 11

```

AXO1_HUMAN
ID   AXO1_HUMAN          STANDARD;          PRT;  1040 AA.
AC   Q02246;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL
DE   GLYCOPROTEIN 1);
GN   TAX1 OR TAG1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=93145965; PubMed=8425542;
RA   Hasler T.H., Rader C., Stoeckli E.T., Zueligg R.A., Sonderegger P.;
RT   "cDNA cloning, structural features, and eucaryotic expression of
RT   human TAG-1/axonin-1.";
RL   Eur. J. Biochem. 211:329-339(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=94140354; PubMed=8307567;
RA   Tsiotra C.P., Karagogeos D., Theodorakis K., Michaelidis M.T.,
RA   Modi W.S., Furley J.A., Jessel M.T., Papamatheakis J.;
RT   "Isolation of the cDNA and chromosomal localization of the gene
RT   (TAX1) encoding the human axonal glycoprotein TAG-1.";
RL   Genomics 18:562-567(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC   AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC   GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC   -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X68274; CAA48335.1; -.
DR   EMBL; X67734; CAA47963.1; -.
DR   PIR; S28830; S28830.
DR   MIM; 190197; -.
DR   INTERPRO; IPR001777; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00041; fn3; 4.
DR   PFAM; PF00047; ig; 6.
KW   Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW   Cell adhesion; Repeat.
FT   SIGNAL      1      28

```

FT CHAIN 29 1012 AXONIN-1.
FT PROPEP 1013 1040 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 54 118 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 148 216 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 313 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 341 402 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 433 495 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 523 594 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 606 612 GLY/PRO-RICH.
FT DOMAIN 611 706 FIBRONECTIN TYPE-III.
FT DOMAIN 714 809 FIBRONECTIN TYPE-III.
FT DOMAIN 816 908 FIBRONECTIN TYPE-III.
FT DOMAIN 917 1003 FIBRONECTIN TYPE-III.
FT SITE 794 796 CELL ATTACHMENT SITE (BY SIMILARITY).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 1012 1012 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1040 AA; 113393 MW; 254E78DD3C28EB6 CRC64;

Query Match 8.5%; Score 582.5; DB 1; Length 1040;
Best Local Similarity 26.7%; Pred. No. 1.2e-24;
Matches 220; Conservative 107; Mismatches 349; Indels 149; Gaps 32;

Qy 79 HRIVLDTSGLFLKVNKSGKNGKSDAGAYVCASNEHEGVSKNEGLKMLR---EDFR 135
| : ||: : : : | | | | : : : | | | | |
Db 183 HFVSQTTGNLYIARTNA-----SDLGNYSLATS-HMDFSTKSVFSKAQLNLAEDTR 235
| : ||: : : : | | | | : : : | | | | |
Qy 136 V-----RPRTVQALGGEMAVLESCPPRGFPPEVSVWRKDKELRIQDMPTLHSDG 187
| : ||: : : : | | | | : : : | | | | |
Db 236 LFAPSIAKARPAETALVGQVTLCEAF-AGNPVPRIKWRKVDGSL----SPQWTT-AEP 289
| : ||: : : : | | | | : : : | | | | |
Qy 188 NLIIDPVRSDSGTYQCVAANMVGERSVNPARLSVFEPKPEQPKDMTVDVGAALVDC 247
| : ||: : : : | | | | : : : | | | | |
Db 290 TLQIPSVSFEDEGTTECAENSKG-RDTVQGRIVQAQPEWLKVISSTEADIGSNLRWGC 348
| : ||: : : : | | | | : : : | | | | |
Qy 248 RVTGDPQPPQITWKRKNEMPVTRAYIAKDN-----GLRIERVQPSDEGEYCYARNPA 301
| : ||: : : : | | | | : : : | | | | |
Db 349 AAAGKPRPTVRLRNGEP-----LASQNRVEVLADGLRFSKLSLEDSGMTQCAVENKH 401
| : ||: : : : | | | | : : : | | | | |
Qy 302 GTLEASAHRLVQA-PPSFQTKPADQSVPA--GGTATFECTLVGQSPAYFWSKEGQDLL 358
| : ||: : : : | | | | : : : | | | | |
Db 402 GTIYASAEALVQALAPDFRLNPVRRLIPAARGGEILIPCQRAAPKAVLWSK-GTEILV 460
| : ||: : : : | | | | : : : | | | | |
Qy 359 FPSYVSADGRTKVSPGTGLTIEEVQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQ 418
| : ||: : : : | | | | : : : | | | | |
Db 461 NSS-----RVTVPDGTLLIRNISRDESKYTCFAENFMKANSTGILSVRDATK--- 510
| : ||: : : : | | | | : : : | | | | |
Qy 419 KKKSKMGKQKQKNVQSIIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASG 478
| : ||: : : : | | | | : : : | | | | |
Db 511 -----ITLAPSSADINLGDNTLQCHASH 534
| : ||: : : : | | | | : : : | | | | |
Qy 479 KPTPGI--SWLRDGLPIDITD-----SRISQHSI-GSLHIADLKKPDGTGYTICAKNEDG 530
| : ||: : : : | | | | : : : | | | | |
Db 535 DPTMDLTFTWTLDDFPIDFDKPGGHYRRNTVKETIGDLTILNAQLRHGGKYTCMAQ---- 590
| : ||: : : : | | | | : : : | | | | |
Qy 531 ESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPTQPIIVNVITDEVELHWNAPSTSGAGPI 590
| : ||: : : : | | | | : : : | | | | |
Db 591 -----TVDSASKEATVLVRGP---PGPPGVVVRDIDGTTIQLSWSR-GFDNHSPI 638
| : ||: : : : | | | | : : : | | | | |
Qy 591 TGYIIQYSPDLGQTWNPIDYVASTE-----YRIKGLKPSHSMYFVIRANEKIGITPS 645
| : ||: : : : | | | | : : : | | | | |
Db 639 AKYTLQARTPPAGK-WKQVRNPNANIEGNAETAQVLGLTPWMDYEFVRVIASNIGTGEP 697
| : ||: : : : | | | | : : : | | | | |
Qy 646 VSSALVTTSPKPAQVALSDKNKMDMAIEKRLTSEQLIKLEEYKTIINSTAVRLFWKRRKL 705
| : ||: : : : | | | | : : : | | | | |

Db 698 GPSSKIRTREAPSVAPSLSGGGGAPGE-----LI-----VNWTPMSREYQNG-- 741
| : ||: : : : | | | | : : : | | | | |
Qy 706 ELID--GYIKWRGPPRTNDNQYVNVSTSPSTENYVSN--LMPFTNYEFVPIYHSGVH 761
| : ||: : : : | | | | : : : | | | | |
Db 742 ----DGFGYLLSFRQGSTH-WQARVPGADAQYFVYSNESVRPYTPFEVKIRSYN---R 793
| : ||: : : : | | | | : : : | | | | |
Qy 762 SIHGAPSNMVDVLTAEAPSLPPEDVRIRMLNLTTLRISWKAPKADGINGILKGFQIVI- 820
| : ||: : : : | | | | : : : | | | | |
Db 794 RGDGPESLTALVYSAEERPAVPTKVWAKGVSSSEMNVWE-PVQDDMNGILLGYEIRY 852
| : ||: : : : | | | | : : : | | | | |
Qy 821 -VGQAPNNRNITTNERAASVTLFHLVTGMTYKIRVAARSNGGVG 864
| : ||: : : : | | | | : : : | | | | |
Db 853 KAGDKEAADRVRTAGLDTARSVGLHPNTKYHVTVRAYNRAGTG 897
| : ||: : : : | | | | : : : | | | | |

RESULT 12

CAML_MOUSE

ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-BRAIN;
RX MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplov D., Frueh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
superfamily with binding domains similar to fibronectin.";
RL Nature 334:701-703(1988).
CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
TO AXONIN ON NEURONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X12875; CAA31368.1; -.
DR PIR; S05479; S05479.
DR HSP; P20241; JCFB.
DR MGD; MGI:96721; L1CAM.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
DR PRINTS; PR00014; FNTYPEIII.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Signal; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1124 1146 POTENTIAL.
FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.

FT	DOMAIN	346	410	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	440	503	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	531	599	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	827	896	FIBRONECTIN TYPE-III.
FT	DOMAIN	932	994	FIBRONECTIN TYPE-III.
FT	DOMAIN	1032	1094	FIBRONECTIN TYPE-III.
FT	SITE	553	555	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	562	564	CELL ATTACHMENT SITE (POTENTIAL).
FT	CARBOHYD	100	100	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	202	202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	246	246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	293	293	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	478	478	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	489	489	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	587	587	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	725	725	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	776	776	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	824	824	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	848	848	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	875	875	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	968	968	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	978	978	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1022	1022	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1107	1107	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1180	1183	MISSING (IN SHORT ISOFORM)
FT				(BY SIMILARITY).
SO	SEQUENCE	1260	AA: 140968	MW: 22B57D01CB2A538 CRC64;

Query Match 8.3%; Score 567.5; DB 1; Length 1260;
Best Local Similarity 21.8%; Pred. No. 1e-23;
Matches 271; Conservative 173; Mismatches 470; Indels 327; Gaps 51;

```

Qy      30 PVIIIEH-PIDVVSRGSPATLNCAG-PSTAKITWYKDG---QPVITNKBEQVNSHRVLD 84
      ||| | | :|| :|| | : ||| | : | || : : :|
Db      35 PVITEQSPRRLVFPFTDDISLCKEARGRPQVEFRWTKDGIHFKP---KEELG---VVVH 87

Qy      85 ---TGSFLLLKVNKGNGKDKSDA---GAYYCVASNEHEGVKSNEGSLKLAMLREDFRV 136
      :|| : : ||| | : ||| | : ||| | : ||| : :|
Db      88 EAPISGSGFTI-----EGNNSFAQRFGIYRCYASNKLGTAHSHE---IQLVAEGAPK 136

Qy     137 RP---RTVQALGGEMAVLECSPPRGFPPEPVSVWRKDDKELRIQDMPRYTLHSDGNLIID 192
      :|| : | | ||| | : ||| | : ||| | : ||| :|
Db     137 WPKETVKPVEVEEGESVVLPCNPPPSAAPPRIYW-MNSKIPDIQDERVSMQNGDLYFA 195

Qy     193 PVDRSDS-----GTYYQCVANNMVGVERSVNPARLSVFEEKP--FEQEPKDMTVDV 239
      | ||| : : ||| : : ||| : : ||| : : ||| :|
Db     196 NVLTSDNHSDYICNAHFPGRTTIIQKEPIDLRV-KPTNSMIDRKPLRLFTNSSSRLVAL 254

Qy     240 -GAAVLFCDCRVTDGPQPIQTKWRKKNEMPVTRAYIAKDNRGLRIERVQPSDEGEYVCYAR 298
      | : : : | : ||| | : ||| | : ||| : | : ||| |
Db     255 QGQSLLILECIAEGFPTPTIKWLHPSPMPTRDVIYQHNKTLQLLVNGEEDDGEYTCIAE 314

Qy     299 NPAGTLEASAHLRVQAPPSFQTKPADQSVVAGGTATFECTLVGQPSPAYFWSKEGQDLL 358
      | | : : : ||| : | : ||| : | : ||| : | : ||| :|
Db     315 NSLGSARHAYYVTEAAPYWLKRPQSHLYGPGETARLDCVQGRPQPEITWRING----M 370

Qy     359 FPSYVSADGRTKVSPGTGLTIEEVQRVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQ 418
      | : | : : ||| : | : ||| : | : ||| : | : ||| :|
Db     371 SMETVKNQDKYRIE-QGSLILSNVQPTDTMTVTCQEARNQHGILLANAYIYV-QLPARIL 428

Qy     419 KKKSKMGKQKQKNVQSIYKILISAVTGNTPAKPPTTIEHGHNQNTLMV--GSSAILPCQA 476
      | : ||| | ||| | ||| | ||| | ||| | ||| |
Db     429 TK-----DNTYTMAVEGSTAYLLCKA 449

Qy     477 SGKPTPGISWLRLGLPIDITSRIQSHTGSLHIADLKKPDTGYVTICAKNEDGESTWSA 536
      | : | : : | : ||| : ||| | ||| | ||| | ||| |
Db     450 FGAPVPSVOWLDEEGTTVLDERFFPYANGTSLIRDLOANDTGRYFCOAANDONNVTLIA 509

```

```
Qy 537 SLTVEHDSNAQFVRMP-----DPS----- 556
   :|::|||
Db 510 NLQVKEAQTITQGPRSAIEKKGARVTFTQCASFDPSLQASITWRGDGRDLQERGSDSKYF 569

Qy 557 -----NFP-----SSPTQPIIVN-----VTDTE 574
   |:|:|:|
Db 570 IEDGKLVISLDYSDOGNYSVCVASTELDEVESRAQLLWVGSPGPVPHLESLDRHLLKQSQ 629

Qy 575 VELHWNAPSTSGAGPITGYIIQQYSPDLG-QTFWN---IPDYVASTEYRIKGLKPSHSYM 630
   |||:|:|||:|:|:|:|:|:|:|:|:|:|
Db 630 VHLNW-SPAEDHNSPIEKYDIEFEDKEMAPEKWFSLGKVPGNQTSITLK---LSPYVHYT 685

Qy 631 FVIRAENEGKTGPS-VSSALVTTSKPAAQVALSDKNKMDMAIAEKLRTSEQLIKLEEVS 689
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 686 FRVTAINKYGPGEPSPVSESVVTP-----AAEPKNPDVR-GEGNETNNMVIWKPLR 738

Qy 690 TINSTAVRLFVKRRKLEELIDYYIKWRGPRTNDNQYNVNTPSTENYVSNLMPTNY 749
   ::||:|:|:|:|:|:|:|:|:|:|:|:|
Db 739 WMD-----HNAPQIQ-----YRVQRPPQKGQETWRKQTVSDPF---LVVNSTSTVFVPY 783

Qy 750 EFFVIPYHSGVHSHGAPSNSMDV-LTAEAPPSPPEDEVIRIMLNLTLRISWKAKPADG 808
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 784 EIKV---QAVNNQKGKPEPQVTIGYSGEDTPQVSPLEDIIFNSSVLVWRVRVDLAQ 839

Qy 809 INGILKGFQIVI---VGQAPNNNRNITTNE-----RAASVTLFHLVTGMTYKIRVAARSN 860
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 840 VKGLHGKYNVYVWWKGSQRKHKSRIHKSHIVVPANTTSAILSLGLRPYSSSHVEVQAFNG 899

Qy 861 GGVGVSH-----GTSEVIMNQDTLEKHLAAQGENSFYGLINKSHVPVIVIVA 909
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 900 RGLGPASEWTFSTPEGVPGHPAL-----HLECQSDTSLLLHWQPPLSHNGVLT--- 948

Qy 910 ILIIFVVIIIAYCYWRNSRNSDGKRSFIKINDGSVHMASNNLWDVAQNPNQNPMYNTAG 969
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 949 -----GYLLSYHPVEGESKEQLFFNLSD--PELRTHNL-----TNLNP----- 984

Qy 970 RMTMNRNNGOALYSLTPAQDFFNCCDDYSGTMHRPGSEHHYHAQLTGGPGNA----- 1023
   |:|:|:|:|:|:|:|:|:|:|:|:|
Db 985 -----DLQIRFQLQATTQ-----GGPGEAIVREGG 1010

Qy 1024 -MSTF---YGNQYHDDPSYATTTVLVLSNOQPWLNDKMLRAPAMPTNPVPP--EPPAR 1076
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 1011 TMALFGKPPDFGNISATAGENYSVVSWPVRKQGCNERFHILFK--ALPECKVSPDHQPOQ 1068

Qy 1077 YADTAGRRSRSSRASDGRGTLN-----GGLHH---RTSGS 1109
   |:|:|:|:|:|:|:|:|:|:|:|:|
Db 1069 YYSYNQSSYTOWNLOPDTKYEHLIKEVLLHHLDVKTNGT 1109
```

RESULT 13

```

AX01_CHICK
ID  AX01_CHICK      STANDARD;      PRT;   1036 AA.
AC  P28685;
DT  01-DEC-1992 (Rel. 24, Created)
DT  01-DEC-1992 (Rel. 24, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  AXONIN-1 PRECURSOR.
    Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  [1]
RP  SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC  TISSUE=BRAIN;
RX  MEDLINE=92174898; PubMed=1311675;
RA  Zuellig R.A., Rader C., Schroeder A., Kalousek M.B.,
RA  von Bohlen Und Halbach F., Osterwalder T., Inan C., Stoeckli E.T.,
RA  Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
RT  "The axonally secreted cell adhesion molecule, axonin-1. Primary
RT  structure, immunoglobulin-like and fibronectin-type-III-like domains
RT  and glycosyl-phosphatidylinositol anchorage.";
RL  Eur. J. Biochem. 204:453-463(1992).

```

CC -!- FUNCTION: AXON-ASSOCIATED CELL ADHESION MOLECULE (AXCAM) WHICH
CC PROMOTES NEURITE OUTGROWTH BY INTERACTION WITH THE AXCAM L1 (64)
CC OF NEURITIC MEMBRANE.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR.
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X63101; CAA44815.1; -.
DR PIR; S22128; S22128.
DR PIR; S22383; S22383.
DR HSP5; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.

FT	SIGNAL	1	23	OR 25 (POTENTIAL).
FT	CHAIN	24	1036	AXONIN-1.
FT	PROPEP	?	1036	REMOVED IN MATURE FORM.
FT	DOMAIN	49	113	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	143	211	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	249	308	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	336	397	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	428	490	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	518	589	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	599	608	HINGE (POTENTIAL).
FT	DOMAIN	601	607	GLY/PRO-RICH.
FT	DOMAIN	608	709	FIBRONECTIN TYPE-III.
FT	DOMAIN	710	811	FIBRONECTIN TYPE-III.
FT	DOMAIN	812	912	FIBRONECTIN TYPE-III.
FT	DOMAIN	913	1009	FIBRONECTIN TYPE-III.
FT	MOD_RES	724	724	BLOCKED.
FT	CARBOHYD	71	71	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	199	199	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	456	456	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	472	472	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	493	493	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	900	900	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	914	914	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1036 AA;	113301 MW;	08B80143BE779794 CRC64;

Query Match 8.2%; Score 562.5; DB 1; Length 1036;
Best Local Similarity 23.3%; Pred. No. 1.5e-23;
Matches 229; Conservative 113; Mismatches 371; Indels 271; Gaps 34;

Qy	30	PVIEHPIDVVSRGS---PATLNCGAK---PSTAKITWYKDGQPVITNKEQVNSHRIVL	83
Db	32	PVFEEQPAHTLFEPSAEKVTILTCRARANPPATYR--WKMG---TELMGPDSTRYL	85
Qy	84	DTGSLFLLKVNCGKSDAGAYCYVASNEHGEVKSNEGSLKLAHLRE-----	132
Db	86	VAGDLVI-----SNPVKAKDAGSYQCVATNARGTVVVSREASLRFGLQFEFSAERDPVKI	140
Qy	133	-----	132
Db	141	TEGWGMFTCSPPHPALSYRWLLNEFPNFIADGRRFVSQTTGNLYIAKTEASDLGNY	200
Qy	133	-----DFRVR-----PRTVQALGGEMAVLECSPPR	157
Db	201	SCFATSHIDFITKSVFSKFSQLSLAEDARQYAPSIKAKFPADTYALTQGMVTECF-AF	259

Qy	158	GFPEPVVSRKDDKELRIQDMPRYTLHSDGNLIIDPVDRSDSGTYQCVANMVGERVSNP	217
Db	260	GNPVPQIKWRKLD-----GSQTSKWLSEPLLIHQNVDFEDEGTYECEANIKG-RDTYQ	313
Qy	218	ARLSVFEPKPFQEPKDMTVDVGAALFDCRVTDGPQITWKRKNEMPVTRAYIAKDN	277
Db	314	GRIIHAQPDWLDVITDTEADIGSDLRWSCVASGKPRPAVRWLRDQGP-----LASQN	366
Qy	278	R-----GLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQA-PPSFQTKPADQSVPA-	329
Db	367	RIEVSGGELRFSKLVLEDSGMTQCVAKENKGTIVYASAEITVQALAPDFRLNPVKRLIPAA	426
Qy	330	-GGTATFECTLVGQSPAYFWSKEGQDLFFPSYVSADGRKTVSPTGTLTIEVRQVDEG	388
Db	427	RSQKVIIPQCPRAAPKATVLWTKGTE-----LLTNSSRVTITADGTLILQNIKSDEG	479
Qy	389	AYVCAGMNSAGSLSKALKATFETKGRVQKKSKMGKQKQNVQSIKYILISAVTGNTP	448
Db	480	KYTCAENFMGKANSTGILSVRDATK-----	505
Qy	449	AKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGI--SWLRDGLPIDITDS-----RIS	501
Db	506	-----ITLAPSSADINVENLTQCHASHDPTMDLFTWSLDFFIDLDKSEGHYRRAS	559
Qy	502	-QHSTGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFP	560
Db	560	VKEAVGDLAIVNAQLKHSRGYICTAQ-----TVVDSTSESATLTVRGP---PG	604
Qy	561	SPTQPIIVNVITTEVELHWNAPSTSGAGPTGTYIIQYSPDLGQTFNIPDYASTE---	617
Db	605	PPGVVVRDIDGDTTVQLSWSR-GFDNHSPIARYSTEARLT-LSNKKQMRTNPVNIENGA	662
Qy	618	--YRIKGLKPSHSMFVIRAENKEGIGTPSVSSALVTTSKPAAQVALSDKNMDMAIAEK	675
Db	663	ETAQVNNLIPWMDYEFVRVLSNGLVGEPSPSSKIRTKEAAPTIVAPS-----GLGGG	715
Qy	676	RLTSEQLIKLEEVKTIINSTAVRLFWRKKRLEELID-----GYIYKWRGPRTNDNQYV	728
Db	716	GGAPNELI-----INWPT-----LRDYQNGDGFYILSFR---KKGTOGWL	754
Qy	729	NVTSPSTE--NYVVSN--LMPFTNYEFFVPIYHSGVSHIGAPSNSMDVLTAEAPPSLPP	784
Db	755	TARVPHASLHYVYRNESIGPTPFVVKIKAYN---RKGEPESLTAIVYSAEPEPKVAP	811
Qy	785	EDVRIRMLNLTTLRISWKAPKADGINGILKGFQIVIVGQAPNNRNNTTNERAASVTLFH	844
Db	812	FRVTAKAVLSSEMDVSWEPVEQDMTGVLGLYIERY--WKDGKBEAADRVRTAGLVTSA	869
Qy	845	LVTGMT---YKIRVAARSNGGVC	864
Db	870	HVTGLNPNTKYHVSVRAYNRAGAG	893

RESULT 14

NRCA_CHK

ID	NRCA_CHK	STANDARD;	PRT;	1284 AA.
AC	P35331;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	NG-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
RN	[1]			
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.			
RP	STRAIN-WHITE LEHORN; TISSUE=EMBRYONIC BRAIN;			
RX	MEDLINE=91258407; PubMed=2045418;			
RA	Grumet M., Mauro V., Burgoon M.P., Edelman G.M., Cunningham B.A.;			
RT	"Structure of a new nervous system glycoprotein, Nr-CAM, and its			
RT	relationship to subgroups of neural cell adhesion molecules."			

```

J. Cell Biol. 113:1399-1412(1991).
[2]
SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.
TISSUE=EMBRYONIC BRAIN, AND RETINA;
MEDLINE=92381110; PubMed=1512296;
Kayyem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;
"Bravo/Nr-CAM is closely related to the cell adhesion molecules L1
RT and Ng-CAM and has a similar heterodimer structure.";
J. Cell Biol. 118:1259-1270(1992).
-1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
NEURITES, ETC. SPECIFICALLY INVOLVED IN THE DEVELOPMENT OF OPTIC
FIBRES IN THE RETINA.
-1- SUBUNIT: HETERODIMER, COMPOSED OF AN ALPHA AND A BETA CHAIN.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS ARE PRODUCED BY
ALTERNATIVE SPLICING.
-1- TISSUE SPECIFICITY: RETINA AND DEVELOPING BRAIN.
-1- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING NEURAL RETINA AND
EMBRYONIC BRAIN TISSUE.
-1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
-1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X58482; CAA41391.1; -.
EMBL; L08960; AAA48632.1; -.
HSSP; P20241; 1CFB.
INTERPRO; IPR001777; -.
INTERPRO; IPR003006; -.
PFAM; PF00041; fn3; 5.
PFAM; PF00047; ig; 6.
PRINTS; PR00014; FNTYPEIIII.
Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
Transmembrane; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 1284 NG-CAM RELATED CELL ADHESION MOLECULE.
FT DOMAIN 25 1143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1144 1166 POTENTIAL.
FT DOMAIN 1167 1284 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 56 125 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 155 220 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 261 323 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 351 415 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 445 508 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 599 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 638 699 FIBRONECTIN TYPE-III.
FT DOMAIN 738 799 FIBRONECTIN TYPE-III.
FT DOMAIN 837 906 FIBRONECTIN TYPE-III.
FT DOMAIN 943 1006 FIBRONECTIN TYPE-III.
FT DOMAIN 1057 1114 FIBRONECTIN TYPE-III.
FT DISULFID 63 118 POTENTIAL.
FT DISULFID 162 213 POTENTIAL.
FT DISULFID 268 316 POTENTIAL.
FT DISULFID 358 408 POTENTIAL.
FT DISULFID 452 501 POTENTIAL.
FT DISULFID 543 592 POTENTIAL.
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 995 995 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1059 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 612 621 MISSING (IN ISOFORM AS10).
FT VARSPLIC 1027 1038 MISSING (IN ISOFORM AS12).
FT VARSPLIC 1039 1131 MISSING (IN ISOFORM AS93).
FT VARSPLIC 1202 1205 MISSING (IN ISOFORM AS-CYT2).
FT CONFLICT 209 209 V -> E (IN REF. 2).
FT CONFLICT 680 680 H -> Q (IN REF. 2).
SO SEQUENCE 1284:AA: 141851 MW: A3570BF9C3D47A0F CRC64:

```

Query Match 8.2%; Score 561; DB 1; Length 1284;
Best Local Similarity 22.1%; Pred. No. 2.4e-23;
Matches 214; Conservative 138; Mismatches 389; Indels 226; Gaps 33;

```

Qy 30 PVIIIEH-PIDVVSRGSPATLNCGAK---PSTAKITWYKDGQPVITNKQVNSHRIVLD 84
   ||: ||:| :||| || :|||
Db 41 PTITQQSPKDYIVDPRENIVIQCEAKGKPPPS---FSWTRNGT-----HFDID 85

Qy 85 TGSFLFLKVNSG-----KNGKSDA--GAYICVASNEHGEVKSNEGSLK--LAMLREDF 134
   :||| ||: ||:| ||| ||| ||: ||
Db 86 KDAQVTMPKNSGTLVVNIMNGVKAEATYEQCTARNERGAASNNVIRPSRSPLWTKE 145

Qy 135 RVRPRTVQALGGEMAVLECSPPRGFPPEVPSWRKDDKELRIQDMPTRYTLHSDGNLIIDPV 194
   :||: ||:| ||: ||:| ||: ||: ||: ||
Db 146 KLEPNHVRE--GDSLVLNCRPPVGLPPPIIFW-MDNAFQLRQPSERVSQGLNGDLYFSNV 202

Qy 195 DRSDSGT-YQCVAA--NNMVGERVSNPARLSVFE-KPKFEQEPKDMT-----VDV-GA 241
   :||: ||: ||: ||: ||: ||: ||: ||
Db 203 QPEDTRVDYICYARFNHTQTIQQKQPISVKVFSTKPVTERPPVLLTPMGSTSNKVELRGN 262

Qy 242 AVLDFCRVTDGDPQQTITWKRKNPEMPMPTRAYIAKDNRLGRIERVQPSDEGEYVCYARNPA 301
   :||: ||: ||: ||: ||: ||: ||: ||
Db 263 VLLLECTIAAGLPTPVIRWIKEGGELPANRTFFENFKKTLAIDVSEADSGNYKCTARNVL 322

Qy 302 GTLEASALRLVQAPPSFQTKPADQSVPAAGTATFECTLVGQSPSPAYFWSKEGQQDLFFPS 361
   ||: ||: ||: ||: ||: ||: ||: ||
Db 323 GSTHHVIVSTVKAAPYWITAPRNLVLSPGEDGTLICRANGNPKPSISWLTVG-----VPI 377

Qy 362 YVSADGRTKVSPGTGLTIEEVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKK 421
   :||: ||: ||: ||: ||: ||: ||
Db 378 AIAPEDPSSRKVDGDTIIFSAVQERSSAVQCNASNEYG----- 415

Qy 422 SKMGKQKQKNVQSIILKYILSAVTGNTPAKPPTTIEHGHNQTLTMVGSSAILPCQASGKPT 481
   ||: ||: ||: ||: ||: ||: ||: ||
Db 416 -----YLLANAFVNLAEPRILTPANKLYQVIADSPALIDCAIFYGSPK 459

Qy 482 PGISWLRDGLPIDITDSRISQHTGSLHIADLKKPDTGVYTICAKNEDGESTWSASLTVE 541
   ||||| :||: ||: ||: ||: ||: ||: ||
Db 460 PEIEWFRGVKGSILRGNIEYVFDNGLTLEIPVAQDKSTGTITVCARNKLKGTQNEVQLEVY 519

Qy 542 DHT-----SNAQFV-----RMPDPNSFPSPSTPII 567
   || :||| :|||
Db 520 DPTMIIKQPYKVIQRSQAQSFCEVIKHDPTLIPTVWLNKDNBELPDERFLVGKDNLTI 579

Qy 568 VNVTDTE-----V 575
   :||| :
Db 580 MNVTDKDDGTTCIVNTTLDVSASAVLTVAAAPPTPAIIYARNPNPLDLTLTGQLERSI 639

Qy 576 ELHWNAPSTSGAGPITGYIIQY----YSPDLGQTFWNPIDYVASTEYRIKGLKPSHSYMF 631
   ||| ||: ||: ||: ||: ||: ||: ||
Db 640 ELSW-VPGEENNSPITNFVIEYEDGLHEPGVWHYQTEVPG--SHTVQLK-LSPYVNYSF 695

Qy 632 VIRAEENERGICTPVSVALVTTSKPAQVALSDKNKMDMAIEAKRLTSEQ---LIKLEEV 688
   :||| ||: ||: ||: ||: ||: ||
Db 696 RVIAVNEIGRSOP-----SEPSEQYLTKSANDPENSPNVQIGTSEPNDLVITWESL 746

```

```

Qy      30 PVIIIE-PIDVVVSRGSPATLNCGAK-PSTAKITWYKDG---QPVIITKEQVNSHRIVLD 84
      ||| | | :|| :|| | : : ||| | : ||| :| :
Db      35 PVITEQSPRRLVVFPTDDISLKECEARGRPQVEFRWTKDGIHFKP----KEELG---VVVH 87

Qy      85 ----TGSFLFLKVNSGKNGKSDA---GAYYCVASNEHGEVKSNEGSLKLAMREDFRV 136
      :|| : | :|| :|| | :|| | :|| :| :|
Db      88 EAPYSGSFTI-----EGNNSFAQRFQGIYRCYASNLTGAMSHE----IQLVAEGAPK 136

Qy     137 RP----RTVQALGGEMAVLECSPPRGFPPEVVSWRKDDKELRIQDMPRYTLHSDGNLIID 192
      | :| :| :| ||| :|| :| :| ||| :| :| :|
Db     137 WKETVKPVEYEEGESVVLPCNPPPSAAPLRIYW-MNSKILHIKQDERVSMGQNGDLYFA 195

Qy     193 PVDRSDS-----GTQYCVANNMVERVSNPARLSVFKEPKF---EQEPKDMTVD 238
      | ||| :|| :| :| :| :| :| :| :|
Db     196 NVLTSDNHSYLCNAHFPGTRTIIQKEPIDLRV-KPTNSMIDRKPRLLFPTNSSSHLVAL 254

Qy     239 VGAAVLFCDCRWGDPQPIQITWKRKNEMPMPVTRAIYAKDNRLRIERVQPSDEGEYVCYAR 298
      | :| :| :| :|| :||| :| :| :| :| :| :|
Db     255 QQGSLLILECIAEGFPTPTIKWLHSPDMPMTDRVYIQHNHNTQLLNVGEEDDGeyTCLAE 314

Qy     299 NPAGTLEASAHLEVRQAPPSFQTKPADQSVPAAGGTATFECTLVGQPSPAYFWSKEGOQDLL 358
      | | : : :| :| :| :| :| :| :| :| :| :|
Db     315 NSLGSARAHYIVTVEAAPYWLQKPQSHLYGPGETARLDCQVQGRPQPEVTVWRING----M 370

Qy     359 FPSYVSADGRKTVSPGTGLTIEEVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQ 418
      | :| :| :| :| :| :| :| :| :| :| :|
Db     371 SIEKVNKQDKYRIE-QGSLLISNVQPSDTMVTQCEARNQHGLLLANAYIVV-QLPARIL 428

Qy     419 KKKSKMGKQKQKNVQSIKIYILSAVTGNTPAKPPTTIEHGHNQNTLMV--GSSAILPCQA 476
      | :| :| :| :| :| :| :| :| :| :|
Db     429 TK-----DNQTYMAVEGSTAYILLCKA 449

Qy     477 SGKTPGISWLRDGLPIDITDSRISQHSSTGLSHIADLKKPDTGVYTICIAKNEDGESTWSA 536
      ||| :| :| :| :| :| :| :| :| :| :|
Db     450 FGAPVPSVOWLDEEGTIVLODERFFPYANGHIGIRDLOANDTGRYFCOAANDONNVITILA 599

```

[illegible]

Search completed: January 22, 2001, 12:28:41
Job time: 1222 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:52:00 ; Search time 559.88 Seconds
(without alignments)
271.520 Million cell updates/sec

Title: US-09-540-245A-17
Perfect score: 6860
Sequence: 1 MYLGFYHTHTHTHTYINFD.....TAQRFRSIPRNGIVTQEQT 1297

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6523.5	95.1	1273	5	O44928	O44928 caenorhabdi
2	4628	67.5	874	5	O01632	O01632 caenorhabdi
3	2232	32.5	423	5	P91572	P91572 caenorhabdi
4	1588	23.1	1395	5	O44924	O44924 drosophila
5	1587	23.1	1395	5	Q9W213	Q9W213 drosophila
6	1505.5	21.9	1612	11	O89026	O89026 mus musculu
7	1500.5	21.9	1651	4	Q9Y6N7	Q9Y6N7 homo sapien
8	1483.5	21.6	1651	11	O55005	O55005 rattus norv
9	1385	20.2	1060	11	Q9Q2I3	Q9Q2I3 rattus norv
10	1323	19.3	1344	11	Q9Z2I4	Q9Z2I4 mus musculu
11	1249	18.2	859	5	Q9VP26	Q9VP26 drosophila
12	1120.5	16.3	823	5	Q9VQ10	Q9VQ10 drosophila
13	674	9.8	1493	11	P97798	P97798 mus musculu
14	658.5	9.6	2016	5	Q9V4J9	Q9V4J9 drosophila
15	658.5	9.6	2016	5	Q9NBA1	Q9NBA1 drosophila
16	642	9.4	1377	11	P97603	P97603 rattus norv
17	640.5	9.3	1461	4	O00340	O00340 homo sapien
18	639.5	9.3	1461	4	Q9Z859	Q9Z859 homo sapien
19	635	9.3	1028	11	Q62682	Q62682 rattus norv

20	632.5	9.2	1443	13	Q90610	Q90610 gallus gall
21	616	9.0	1028	11	Q07409	Q07409 mus musculu
22	609	8.9	1099	11	P97527	P97527 rattus norv
23	608.5	8.9	1026	11	Q62845	Q62845 rattus norv
24	608.5	8.9	1100	4	O94779	O94779 homo sapien
25	607.5	8.9	1040	13	Q9W675	Q9W675 brachydanio
26	607	8.8	1375	5	Q94537	Q94537 drosophila
27	606	8.8	1026	4	O94780	O94780 homo sapien
28	601.5	8.8	1252	11	Q9JLI1	Q9JLI1 mus musculu
29	596	8.7	1028	11	Q9JMB8	Q9JMB8 mus musculu
30	595	8.7	1021	11	Q63198	Q63198 rattus norv
31	592	8.6	1018	6	Q28106	Q28106 bos taurus
32	589	8.6	1272	13	Q90924	Q90924 gallus gall
33	588	8.6	2222	5	O97394	O97394 drosophila
34	587.5	8.6	2221	5	Q9ULM1	Q9ULM1 drosophila
35	586	8.5	1028	11	P97528	P97528 rattus norv
36	583	8.5	1028	4	Q9UQ52	Q9UQ52 homo sapien
37	580.5	8.5	1240	4	O14631	O14631 homo sapien
38	580.5	8.5	1427	13	Q91562	Q91562 xenopus lae
39	577	8.4	1248	6	Q9XT41	Q9XT41 cercopithec
40	572	8.3	1369	13	O42414	O42414 gallus gall
41	565	8.2	1009	13	O93250	O93250 xenopus lae
42	564	8.2	1259	11	Q9QY38	Q9QY38 mus musculu
43	562	8.2	920	4	Q9P232	Q9P232 homo sapien
44	562	8.2	1250	11	O88971	O88971 mus musculu
45	557	8.1	1445	11	Q63155	Q63155 rattus norv

ALIGNMENTS

RESULT 1
O44928
ID O44928 PRELIMINARY; PRT; 1273 AA.
AC O44928;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE SAX-3.
GN SAX-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117250; PubMed=9458046;
RA Zallen J.A., Yi B.A., Bargmann C.I.;
RT "The conserved immunoglobulin superfamily member SAX-3/Robo directs
RT multiple aspects of axon guidance in C. elegans.";
RL Cell 92:217-227(1998).
DR EMBL; AF041053; AAC38848.1; -.
DR HSSP; P56276; IFLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 1273 AA; 139427 MW; 013E766B51A7BAD7 CRC64;

Query Match 95.1%; Score 6523.5; DB 5; Length 1273;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1243; Conservative 1; Mismatches 1; Indels 33; Gaps 3;

Qy 24 NASNLAPVIEHPIDVVVSRGSPATLNCGAKPSTAKITWYKDGQPVITNKEQVNSHRIVL 83
:|||||
Db 25 DASNLAPVIEHPIDVVVSRGSPATLNCGAKPSTAKITWYKDGQPVITNKEQVNSHRIVL 84
Qy 84 DTGSLFLKLVNSGKNGKSDAGAYYCVASNEHGEVKSNEGSLKLAHLREDFRVRPRTVQA 143
:|||||
Db 85 DTGSLFLKLVNSGKNGKSDAGAYYCVASNEHGEVKSNEGSLKLAHLREDFRVRPRTVQA 144

Qy 144 LGGEMAVLECSPPRGFPPEVSVWRKDDKELRIQDMPTLHSDGNLIIDPVRSDSGTYQ 203
Db 145 LGGEMAVLECSPPRGFPPEVSVWRKDDKELRIQDMPTLHSDGNLIIDPVRSDSGTYQ 204
Qy 204 CVANNMVGERSVSNPARLSVFEEKPKFEQPKDMTVDVGAAVLFCRVTGDPQQTWKRN 263
Db 205 CVANNMVGERSVSNPARLSVFEEKPKFEQPKDMTVDVGAAVLFCRVTGDPQQTWKRN 264
Qy 264 EPMPTVTRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASALRVQAPPSFQTKPA 323
Db 265 EPMPTVTRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASALRVQAPPSFQTKPA 324
Qy 324 DQSVPAAGTATFECTLVGQPSPAYFWSKEGQDLPSPYSADGRTKVSPGTLTIEEVR 383
Db 325 DQSVPAAGTATFECTLVGQPSPAYFWSKEGQDLPSPYSADGRTKVSPGTLTIEEVR 384
Qy 384 QVDEGAYVCAGMNSAGSSLSKAALKATFETKRVQKKSKMGKQKQKNVQSIKYLISAV 443
Db 385 QVDEGAYVCAGMNSAGSSLSKAALKVT--TK-----AV 415
Qy 444 TGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRDGLPIDITDSRISQ 503
Db 416 TGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRDGLPIDITDSRISQ 475
Qy 504 STGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPT 563
Db 476 STGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFPSSPT 535
Qy 564 QPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLGQTFWNPIDYVASTERYIKGL 623
Db 536 QPIIVNVTDEVELHWNAPSTSGAGPITGYIIQYSPDLGQTFWNPIDYVASTERYIKGL 595
Qy 624 KPSSHYMFVIRAENEKGIGTSPVSSALVTTSPKAAQVALSDKNKMDMAIEAKRLTSEQLI 683
Db 596 KPSSHYMFVIRAENEKGIGTSPVSSALVTTSPKAAQVALSDKNKMDMAIEAKRLTSEQLI 655
Qy 684 KLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTNDNQYVNVTSPTENYVVSNL 743
Db 656 KLEEVKTINSTAVRLFVKRRKLEELIDGYIIKWRGPPRTNDNQYVNVTSPTENYVVSNL 715
Qy 744 MPFTNYEFFVPIYHSGVHSIHGAPSNMVDLTAEAPPSLPEDVIRMLNLTTLRISWKA 803
Db 716 MPFTNYEFFVPIYHSGVHSIHGAPSNMVDLTAEAPPSLPEDVIRMLNLTTLRISWKA 775
Qy 804 PKADGINGILKGFQIVIVQAPNNRNITTNERAASVTLFHLVTGMTYKIRVAARSNGGV 863
Db 776 PKADGINGILKGFQIVIVQAPNNRNITTNERAASVTLFHLVTGMTYKIRVAARSNGGV 835
Qy 864 GVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVPIVIVAILIIFVVIIAICY 923
Db 836 GVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVPIVIVAILIIFVVIIAICY 895
Qy 924 WRNSRNSDGKDRSFIKINDGSHVMSNNLWDVAQNPQNPMYNTAGRMTMNNRNGQALYS 983
Db 896 WRNSRNSDGKDRSFIKINDGSHVMSNNLWDVAQNPQNPMYNTAGRMTMNNRNGQALYS 955
Qy 984 LTPNAQDFFNCCDYSMTMRPGSEHHYHIAQLTGGPGNAMSTFYGNQYHDDPSYATTT 1043
Db 956 LTPNAQDFFNCCDYSMTMRPGSEHHYHIAQLTGGPGNAMSTFYGNQYHDDPSYATTT 1015
Qy 1044 LVLSNQQPALNDKMLRAPAMPTNPVPEPPARYADHTAGRRSRSSRASDGRGLNGGLH 1103
Db 1016 LVLSNQQPALNDKMLRAPAMPTNPVPEPPARYADHTAGRRSRSSRASDGRGLNGGLH 1075
Qy 1104 HRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTERRTPPNKTLMDIFIPPPSNPPPP 1163
Db 1076 HRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTERRTPPNKTLMDIFIPPPSNPPPP 1135
Qy 1164 GGHVYD---TATRRQLNRGSTPREDTYDSVSDGAFARVDVNARPTSRNRLGGRPLK 1219
Db 1136 GGHVYDIFQTATRRQLNRGSTPREDTYDSVSDGAFARVDVNARPTSRNRLGGRPLK 1195
Qy 1220 RDDDSQRSSLMDDGGSSEADGENSEGDPVGRGVRKAVPRMGISASTLAHSCYGTNGTA 1279

Db 1196 RDDDSQRSSLMDDGGSSEADGENSEGDPVGRGVRKAVPRMGISASTLAHSCYGTNGTA 1255

Qy 1280 QRFRSIPRNNIGIVTQEQT 1297

Db 1256 QRFRSIPRNNIGIVTQEQT 1273

RESULT 2

O01632

ID O01632 PRELIMINARY; PRT; 874 AA.

AC O01632;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE CODED FOR BY C. ELEGANS CDNA CEESC12R.

GN ZK377.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,

RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Nhan M., Hawkins J.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U88183; AAB52657.1; -.

DR HSP; P56276; iTLK.

DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00041; fn3; 3.

DR PFAM; PF00047; ig; 1.

DR PRINTS; PR00014; FNTYPEIII.

SQ SEQUENCE 874 AA; 95861 MW; BC72270818D734C9 CRC64;

Query Match 67.5%; Score 4628; DB 5; Length 874;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 874; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 424 MGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPG 483

Db 1 MGKQKQKNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPG 60

Qy 484 ISWLRDGLPIDITDSRISQHSSTGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDH 543

Db 61 ISWLRDGLPIDITDSRISQHSGLHIADLKKPDTGVYTCIANKEDGESTWSASLTVEDH 120
Qy 544 TSNAQFVRMPDPSNFPSSPTQPIIVNVDTTEVELHWNAPSTSGAGPITGYIIQYSPDLG 603
|||||
Db 121 TSNAQFVRMPDPSNFPSSPTQPIIVNVDTTEVELHWNAPSTSGAGPITGYIIQYSPDLG 180
Qy 604 QTWNPIDYVASTEYRIKGLKPSHSYMFVIRAEKNGIGTSPVSSALVTTSKPAAQVALS 663
|||||
Db 181 QTWNPIDYVASTEYRIKGLKPSHSYMFVIRAEKNGIGTSPVSSALVTTSKPAAQVALS 240
Qy 664 DKNKMDMAIAEKRLTSEQLIKLEEVTINSTAVRLFVKRRKLEELIDGYIKWRGPPRTN 723
|||||
Db 241 DKNKMDMAIAEKRLTSEQLIKLEEVTINSTAVRLFVKRRKLEELIDGYIKWRGPPRTN 300
Qy 724 DNQYVNVTSPTSTENYVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNMVDLTAEAPPSLP 783
|||||
Db 301 DNQYVNVTSPTSTENYVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNMVDLTAEAPPSLP 360
784 PEDVRIRMLNLTTLRISWKAPKADGINKGFQIVIVGAPNNRNITTNERAASVTLF 843
|||||
Db 361 PEDVRIRMLNLTTLRISWKAPKADGINKGFQIVIVGAPNNRNITTNERAASVTLF 420
Qy 844 HLVGTMTYKIRVAARSNGVGVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVP 903
|||||
Db 421 HLVGTMTYKIRVAARSNGVGVSHGTSEVIMNQDTLEKHLAAQENESFLYGLINKSHVP 480
Qy 904 VIVIVAILIIFVVIAYCYWRNSRNSDGRSFKINDGSVHMASNNLWDVAQNPNQNP 963
|||||
Db 481 VIVIVAILIIFVVIAYCYWRNSRNSDGRSFKINDGSVHMASNNLWDVAQNPNQNP 540
Qy 964 MYNTAGRTMNNRNGQALYSLTPNAQDFNNCCDYSGMTMHRPGSEHHYHYAQLTGGPGNA 1023
|||||
Db 541 MYNTAGRTMNNRNGQALYSLTPNAQDFNNCCDYSGMTMHRPGSEHHYHYAQLTGGPGNA 600
Qy 1024 MSTFYGNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPPPEPPARYADHTAG 1083
|||||
Db 601 MSTFYGNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPPPEPPARYADHTAG 660
Qy 1084 RRSRSSRASDGRGLNGGLHRTSGSQRSDSPPHTDVSVQLHSSDGTGSSKERTGERRT 1143
Db 661 RRSRSSRASDGRGLNGGLHRTSGSQRSDSPPHTDVSVQLHSSDGTGSSKERTGERRT 720
Qy 1144 PPNKTLMDFIPPPSPNPPPGGHVYDTATRRQLNRGSTPREDYDSVSDGAFARVDVNAR 1203
Db 721 PPNKTLMDFIPPPSPNPPPGGHVYDTATRRQLNRGSTPREDYDSVSDGAFARVDVNAR 780
1204 PTSNRNLRGRLKGRDDDSQRSSLMDDGGSEADGENSEGDVPRGGVRKAVPRMGI 1263
Db 781 PTSNRNLRGRLKGRDDDSQRSSLMDDGGSEADGENSEGDVPRGGVRKAVPRMGI 840
Qy 1264 SASLHSCYGTNGTAQRFRSIPRNGIVTQEQT 1297
|||||
Db 841 SASLHSCYGTNGTAQRFRSIPRNGIVTQEQT 874

RESULT 3

P91572

ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN ZK377.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nhan M., Hawkins J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88183; AAB52658.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00047; ig; 4.
SQ SEQUENCE 423 AA; 46544 MW; EB4530DB6D575E5 CRC64;

Query Match 32.5%; Score 2232; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.6e-151;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYLGFYHTHTHTYINFDKIPNASLAPVIEHPIDVVSRGSPATLNCAGKPSTAKI 60
Db 1 MYLGFYHTHTHTYINFDKIPNASLAPVIEHPIDVVSRGSPATLNCAGKPSTAKI 60
Qy 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKYNVSGKNGKSDAGAYYCVASNEHEGVKS 120
Db 61 TWYKDGQPVITNKEQVNSHRIVLDTGSLFLKYNVSGKNGKSDAGAYYCVASNEHEGVKS 120
Qy 121 NEGSLKAMLREDFVRPRTVQALGGEMAVLECSPPRGFPPEVVSWRKDDKELRIQDMPR 180
Db 121 NEGSLKAMLREDFVRPRTVQALGGEMAVLECSPPRGFPPEVVSWRKDDKELRIQDMPR 180
Qy 181 YTLHSDGNLIIDPVDRSDSGTYQCVANNMVGERSVNPRLSVFEKPKFEQPKDMTVDVG 240
Db 181 YTLHSDGNLIIDPVDRSDSGTYQCVANNMVGERSVNPRLSVFEKPKFEQPKDMTVDVG 240
Qy 241 AAVLFDRCRVTDQPPQITWKRKNPMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300
Db 241 AAVLFDRCRVTDQPPQITWKRKNPMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNP 300
Qy 301 AGTLEASAHLRVQAPPSFQTKPADQSVVAGGTATFECTLVGQSPAYFWSKEGQDQLLFP 360
Db 301 AGTLEASAHLRVQAPPSFQTKPADQSVVAGGTATFECTLVGQSPAYFWSKEGQDQLLFP 360
Qy 361 SVVSADGRKYSPTGTLTIEEVQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKK 420
Db 361 SVVSADGRKYSPTGTLTIEEVQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKK 420
Qy 421 KSK 423
Db 421 KSK 423

RESULT 4

O44924

ID 044924 PRELIMINARY; PRT; 1395 AA.
AC 044924;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040989; AAC38849.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0005631; robo.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 1395 AA; 151778 MW; B820E234A5218983 CRC64;

Query Match 23.1%; Score 1588; DB 5; Length 1395;
Best Local Similarity 31.0%; Pred. No. 1.7e-104;
Matches 421; Conservative 195; Mismatches 527; Indels 216; Gaps 41;

Qy 29 APVIIHPIIDVVSRGSPATLNC--GAKPSTAKITWYKDGQPVITNKEQVNSHRIVLDTG 86
: | | | | | : | | : | | | | | | | : | | | | | : | | : | | | | |
Db 55 SPRIIEHTDLVVKKNPATLNCVKEGKPEPT-IEWFKDGEFVSTNEKK--SHRVQFKDG 111
: | | | | | : | | : | | | | | | | : | | | | | : | | : | | | | |
Qy 87 SLFLLKNSGKNGKSDAGAYYCVASNEHGEVKSNEGSLKLAILEDPRVRPTVQALGG 146
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 112 ALFPYRTMQGK--KEQDGEYWCVAKNRVQAVSRHASLQIAVLRDDPRVEPKDTRVARG 169
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 147 EMVLECSPPRGFPPEPVSWRKDD-----KELRIQDMPTYLHSDGNLIIDPDRSDSG 200
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 170 ETALLBCCGPKGIPEPTLIWIKDGVPLDDKAMSGASSRVRVDGNNLISNVEPIDEG 229
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Qy 201 TYQCVANMVGERSVNPASLVSFEKPKFEQPKDMTVDGAALVFCRVTDGDPQQTWK 260
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 230 NYKICIAQLVGTRESSYAKLIVQVQKPYFMKEPKDQVMLYGQATFHCSSVGGDPPPKVLWK 289
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 261 RKNEMPVTRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAPPSFT 320
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 290 KEEGNIPVSARILHDEKSLEISNITPTDEGTIVCEAHNNVQISARASLIHAPPNFTK 349
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 321 KPADQSVPAAGTATFECTLVGQSPAYFWSKEGQDQLLPSTVSADGRTVSPTGLTIE 380
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 350 RPSNKKVGLNGVVQLPCMASGNPPSVFWEKGVSTLMFPN--SSHGRQVVAADGTLQIT 407
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 381 EVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKQKNVQSIKYLI 440
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 408 DVRODEGGYVCSAFSVVDSSTVRVELQVS----- 437
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 441 SAVTGNTPAKPPPTIEHGHNQTLWVGSSAILPCQASGKPTPGISWLRLDGLPIDITDSRI 500
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 438 -----SVDERPPPIIQIGPANQTLPGKSVATLPCRATGNPSPRIKWFHDHGAHV-AGNRY 491
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 501 SQHSTGSLHIADLKKPDTGVYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSNFP 560
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Db 492 SIIQSSRLVDDQLSDSGTYTCTASGERGETSWAATLIVKEKPGSTS-LHRAADPSTYPA 550
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |
Qy 561 SPTQPIIVNVTDEVELHW--NAPSTSGAGPITGYYIIQYSPDLGQTFWNPIDYVAST 618
: | | : | | : | | : | | | | : | | : | | | | | : | | : | | | | |

Db 551 PPGTPKVLNVSRTSISLRWAKSQEKPAGVPIIGTYVEYFSPDLQGTWIVAAHRVGDQV 610
Qy 619 RIKGLKPSHYSYMFVIRAEKNGIGTPSSVALVTTSKPAAQVALSDKNKMDMAIAEKRLT 678
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 611 TISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIADFDASAN----DLSAARTLLT 666
Qy 679 SEQLIKLEEVKTINSTAVRLFVKRKL--BELIDGYIKWRGPPRTNDNQY--VNVISPS 734
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 667 GKS-VELIDASAINASAVRLEWMLHVSADKEYEGLRIHYK-DASVPSAQYHSITVMDAS 724
Qy 735 TENYVVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNMMDVLTAEAPPSPDPVIRMLNL 794
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 725 AESFVGNLKKYKYEFLTPF---FETIEGQPSNSKTALTYEDVPSAPPDNIQIGMYNQ 781
Qy 795 TTLRISWKAPKADGINGILKGFQI-VIVGQAPNNRNITTNERASVTFLHLVTGMTYKI 853
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 782 TAGVVRWTPPPSQHNGNLYGKIEVSAGTMMKVLAMNLTATTSVLLNNLTGAVYSV 841
Qy 854 RVAARSNGGVGSHGTSEVIMQDTLEKHLA-----AQENESFLYLINKSHVP- 903
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 842 RLNSFTKAGDGPYSKISLFMD-PTHVHPRAHPSGTHDRHGQDLTYH--NNGNIPP 898
Qy 904 -----VIVIVAILIIFVIIAYCYWRNSRNSDGRSFIKIND 942
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 899 GDINPTTHKKTDTYLSGPNLWLVLCVLLVLVISAISMVYFKRKHQMTKELGHSVVD 958
Qy 943 GSVHMASNN---LWDVAQNPNQNPMTNAGRMNMRNGQALYSLTNPAQDFFNCCDDY 998
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 959 NEITALNINSKESLW-----IDHHRGWRTADTKDGLSEKSLSHVNSQSNYNSD-- 1011
Qy 999 SGTMRPGSEHHYHAQLTGGPGNAMSTFYG-NQYHDDPSYATTLV----- 1045
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1012 GGT-----DYAEV---DTRNLTFYCNCRKSPDNPTPYATTMIIGTSSSETCTKTT 1058
Qy 1046 -----LSNQQA--WLNDKMLRAPAMPTN-----PVPPE--PPARYAD 1079
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1059 SISADKSGTHSPYSDAFAGQVPAVPVVKSNYLQYVPEPINWSEFLPPPEHPPSSSTYG 1118
Qy 1080 HTAG--RRSRSSRASDGRG-----TLNGGLHRTSGSQR-----DSPHTDVS 1122
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1119 YAQGSPESSRSKSSAGSGISTNQSLNASHSSSSGGFSAGWSPQYAVACPPENVYSN 1178
Qy 1123 VQLHSSDGTGSSKERTGERRTPPNKTLMDFIPPPSPNPPPGGHVYDTATRRQ----- 1175
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1179 PLSAVAGGTQNRQITPTNQHPQLPAY-FATTGPGGAVPP-NHL-PFATQRHAASEYQA 1235
Qy 1176 -LNRGSTPREDTYDS-----VSDGAFARVDVNA---RPTSRNRNL--- 1211
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1236 GLNAARCAQSPACNSCDALATPSPMQPPPPVPVPEGWYQVHPNHPMHSTSNHIIYQC 1295
Qy 1212 GGRPLKGRDDDSQRSSLMMDDDGGSSSEADGENSEGDVP 1250
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 1296 SSECSDHRSRSSHQRQLQLEHGSSARQGGHRRRAP 1334

RESULT 5

Q9W213

ID Q9W213 PRELIMINARY; PRT; 1395 AA.
AC Q9W213;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROBO PROTEIN.
GN ROBO.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;


```

RESULT 7
Q9Y6N7
ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6N7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040990; AAC39575.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; iq; 5.

```

[illegible]

RESULT 9
Q9QZI3
ID Q9QZI3 PRELIMINARY: PRT: 1060 AA.

```

Qy      30 PVIIIEHPDIDVWSRGSPATLNCAGK---PSTAKIT-----WKDGGQPVITNKQEVNSH 139
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      31 PKXVEQPSEVIVSKGKPNTPNWKQKGRPPPTIGKVQRMVKPGWDK-----TKDSDKVTQ 84

Qy      80 RIVLDTGSLFLLKVNSGKNGKSDAGAYYCVASNEHEGVKSNEGSLKLMLREDFRVVRP 139
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      85 GCLLPSGSLFFLRIVHGRRSK-PDEGTYVCVARNYLGEAVSRNASLEVALLRDDFRQNPT 143

Qy     140 TVQALGGEMAVLECSPPRGPEPVPVSWRRDKDELRIQDMPRYTLHSDGNLIIDPVDRSDS 199
      | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     144 DVVVAAGEPAILECQPPRGHPEPTIYKKD--KVRIDEKEERISIRGGKLMISNTRKSDA 201

Qy     200 GTYQCVMANMVGERSVNPARLSVFEEKPFKEQPKDMTVDVGAAVLDFCRVTGDPQPQITW 259
      | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     202 GMYTCVGTNMVGERDSDPAELTVFERPTFLRRPINQVWLEDEPAEFRQVQGDQPQPTVRW 261

Qy     260 KRKNEMPMPVTRAIYAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASAHRLVQAPPSFQ 319
      | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     262 KKDDADLPRGR-YDIKDDYTLRIKKAISADEGTYVCIANENRVKGEASATLTVRAPQPVF 320

Qy     320 TKPADQSPVAGGTATFECTLVGQSPSPAYFWSKEGQDQLLFPSY-VSADGRTKVSPGTGLT 378
      : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     321 VRPRDQIVAGRTVTFPCETKGNPQPAVFWQKEGSONLFFNPQQPQQPNRSRCSVSPTGDLT 380

Qy     379 IEEVRQVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKKSKMGKQKQKNVQSIKY 438
      | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     381 ITNIQRSDAGYIYQCALTVAGISLAKAQL----- 410

Qy     439 LISAVGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRLDGLPIDITDS 498
      | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     411 ---VTDVLTDRPPPIIQLGPNQITLAVDGTALLCKATG-PLPVISWLKEGFTFLGRDP 465

Qy     499 RISQHSTGSLHIADLKKPDTVITCIKNEGDGTSWASLTVEDHTSNAQFVRMPDPSNF 558
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     466 RATIQDQGTQLIKNLRISDTGYTTCVATSSSGETSWSAVLDVTE--SGATISKNYDNDL 523

Qy     559 PSSPTQPIIVNVTDEVELHWNAPSTSGAGPTIGYIIQYISPDLGQTFWNPIDYVASTEY 618
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Qy 16 YINFDKIPNA---SNLAPVIIIEHPIDVVVSRGSPATLNCGA----KPSTAKITWYKDQG 67
| : | : | : | : ||||| ||| | :| : |||:|
Db 24 YLELPSSPGSRVGEDAMPRIIVEOPPDILVYSRGEPATLPCRAEGRPRPN---IEWYKNGA 80

[illegible]

Qy 1080 HTAGRRSSRSASDGRGTLNGLLHRTSGSQRSDSPHTDVSIVQLHSSDGTGSSKERTG 1139
Db 1074 CPGEPEELKGSSE-----LEEWCPVPPEKSH--LVGSSSSGACMVAPA 1115

Qy 1140 ERRTTP-----NKTLMDFIPPPSPNPPPGG--HVDYATIRRLQNRG-----STPREDT 1186
Db 1116 PRDTSPSTSSYGQSTATITPSPDPQPPTDIPHLHQMPRRVPLGSPSSPLSVSQPALSS 1175

Qy 1187 YDS--VSDGAFARVDVNARPT-----SRNRNLGGR---PLKGRDDSDQRSSLM 1231
Db 1176 HDGRPVGLGAGPVLSYHASPSVPSTASSAPGRTRQVTGEMTPLLHGRARIRKKPKAL- 1234

Qy 1232 DDDGGSSEADGENSEGDP 1250
Db 1235 -----PYRREHSPGDL 1246

RESULT 11

Q9VPZ6

ID Q9VPZ6 PRELIMINARY; PRT; 859 AA.
AC Q9VPZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG5423 PROTEIN (FRAGMENT).
GN CG5423.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R.C., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).

DR EMBL; AE003586; AAF51388.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0031328; CG5423.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; lg; 5.
FT NON_TER 1 1
SQ SEQUENCE 859 AA; 93916 MW; 5CFD69D984101BF8 CRC64;

Query Match 18.2%; Score 1249; DB 5; Length 859;
Best Local Similarity 32.1%; Pred. No. 1.3e-80;
Matches 309; Conservative 151; Mismatches 351; Indels 152; Gaps 25;

Qy 30 PVIEHPIDVVSRGSPATLNCAGKPS-TAKITWYKDGQPVITNEQVNSHRVLDTGSL 88
Db 1 PRIVEHPIDTTPRHEPATLNCAGKPSPTIQWYKDGVPL--KILPGSHRITLPAGGL 57

Qy 89 FLKVNKGKNGKSDAGAYYCVASNEHEGVKSNESGLKMLREDFVRPRVTQALGEM 148
Db 58 FFLKVSDDR-----RCVLRDFEPLQNTRIAQGT 89

Qy 149 AVLCESPRGFPPEPVSWRKDDKELRIQDMPRYTLHSDGNLIIDPVDRSDSGTYQCVANN 208
Db 90 ALLECAAPRGPEPTVTWKKGGQKLDLEGSKRVRIVDGNLAIQDARQDEGQYQCIAN 149

Qy 209 MVGERVSNPARLSVFEPKFEQEPKMDTVDVGAALVDFCRVTDGPQPOITWKR--KNEPM 266
Db 150 PVGVRESSLATLKVHVKPYIIRGPHDQTVLEGASVTPPCRVGDDPMPVWLRTASGGNM 209

Qy 267 PVTRAYIAKDNRLRIERVQPSDEGEYVYARNPAGTLEASHLRVQAPPSFQTKPADQS 326
Db 210 PLDRVSVLED-RSLRLERVTIADGEYSCEADNVGGAITAMGLTVYAPKFIQRPASKS 268

Qy 327 VPAGGTATFECTLVGQSPAYFWSKEGQDLLEPSYVSADGRTKVSPS---GTLTIEVR 383
Db 269 VELGADTSPECRAIGNPKPTIFWTIKNNSTLIFPGAPPLDRFHSMTTEGHSILTIRFQ 328

Qy 384 QVDEGAYV-CAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKQKNVQSIKYLISA 442
Db 329 RTDKDLVILCNAMNEVASITSRVQL-----SLDSQED----- 360

Qy 443 VTGNTPAKPPPTIEHGHNQTLMVGSAILPCQASGKPTGISWLRDGLPIDITDSRISQ 502
Db 361 -----RPPPIISGPNVQTLPIKSLATLQCKAIGLPSPTISWYRDGPVQ--PSSKLN 412

Qy 503 HSTGSLHIADL-KKPDITGVYTIKAKNEDGESTWSASLTVEDHTS-NAQFVRMPDPSNFP 560
Db 413 TTSGLDIISDLDRQDDQGLTYCVASSRAGKSTWSGFLRIELPTNPNIKYRAPEQTKCPS 472

Qy 561 SPTQPIIVNVTDEVELHWNAPSTSGAGPTIGYIIQYSPDLGQTFWNPIDYVASTERYI 620
Db 473 APGQPKILNATASALTIVWPTSDKAGASSFLGYSVEMYCTNQSRWTIPIASRLSEPIFTV 532

Qy 621 KGLKPSHSYMFVIRAENEKIGTGPSVSSALVTTSK-----PAAQVALSDKNKMD 669
Db 533 ESLTQGAAYMFIVRAENSLGSPSPSISEPITAGLVGRVGDSESTGTSQLLSD----- 587

Qy 670 MAIAEKRLTSEQLIKLEEVTINSTAVRLFVKRRLEELIDGYIKWGPPTNDNQYVN 729
Db 588 ---VETLLQANDVVELLEANASDSTARLSWDIDS-QQYIEGFYLYAR---ELHSEYKM 640

Qy 730 VT----SPSTENYVVSNLMPFTNYEFFVPIYHSGVHSIHGAPSNSMDVLTAEAPPSLPPE 785
Db 641 VTLLNKGQGLSSCTVPLGAKASTYEFFLVFPYK---SIVGKPSNRRRTLEDVPEAPPY 697

Qy 786 DVIRMLNLTTLRISWKAPKADGI-NGILKGFQIVIGQAPNNN----RNITTNERAASV 840
Db 698 GMEAIQFNRTSVFLKWLPPQPNRTNGILTSYNNVVGKLDVHNTTRIFKNMTIDAAAPTL 757

Qy 841 TLFHLVTGMTYKIRVAARSNGGVG-----VSHGTSEVIMNQDTEKHLA 884

```

RESULT 12
Q9VQ10
ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
CG5481 PROTEIN.
CG5481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003586; AAF51373.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0031341; CG5481.
DR INTERPRO; IPR001412; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.

```

DR PFAM: PF00041; fn3; 1.
DR PFAM: PF00047; ig; 5.
DR PRINTS: PR00014; FNTYPEPIII.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN1.
SQ SEQUENCE 823 AA; 89715 MW; 36FC0B91F36F2F19 CRC64;

Query Match 16.3%; Score 1120.5; DB 5; Length 823;
Best Local Similarity 30.3%; Pred. No. 1.9e-71;
Matches 263; Conservative 129; Mismatches 296; Indels 181; Gaps 18;

[illegible]

RESULT 13

P97798
ID P97798 PRELIMINARY; PRT: 1493 AA.
AC P97798;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE NEOGENIN (NEOGENIN PROTEIN).
GN NEOL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse Neogenin, a DCC-like molecule, has four splice variants and is
RT expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
DR EMBL; Y09535; CAA70727.1; -.
DR HSSP; P02751; 1TTF.
DR MGD; MGI:1097159; Neol.
DR INTERPRO; IPR000531; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; PNTYPEIII.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 9.8%; Score 674; DB 11; Length 1493;
Best Local Similarity 21.8%; Pred. No. 3.8e-39;
Matches 334; Conservative 193; Mismatches 551; Indels 452; Gaps 64;

Qy 36 PIDVVVSRGSPATLNCGA--KPSTAKITWYKDGQPVITNKEQVNSHRIVLDTGSLFLKLV 93
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 70 PVDTLVSRGSSVILNCSAYSEPS-PNIEWKKDG--TFLNLES-DDRQLPLDGLSIFSNV 125
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 94 NSGKNGKSDAGAYCYVASNEH-GEVKSNEGSLKLAILEDVFRVRPTVQALGGEMAVLE 152
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 126 VHSKHNK-PDEGFIQCVATVDNLGTVISRTAKLTVAGLPR-FTSQPEPSSVYVGN SAILN 183
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 153 CSPPRGPPEPVVSWRKDDKELRIQDMPTLHSDGNLIIDPVRSDSGTYQCVANNMVG 212
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 184 CEVNADL-VPFVRWQNRQPLLLDD--RIVKLPSGTLVISNATEGDGLYRCIVSGGPP 240
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 213 RVSNPARLSVFEKPK-----FEQEPKDMTVDVGAALVDFCRVTGDPQPIQITWKRKNEM 266
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 241 KFSDEAKLVLDPEEIVDLVFLMRPSSMMKVTGQSAVLPCVVGSLPAPVVRWMKNEEVL 300
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 267 PVTRA--YIADNRGLRIERVQPSDEGEYVCYARNPAGTLEASAHLRVQAPPSFTKPAD 324
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 301 DTSSGRLVLLAGGLEISDVTEDDAGTYFCIADNGNKTVQAELTVQVPPGFLKQFAN 360
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 325 QSVFAGGTATFECTLVGQSPAYFWSKEGQDQLFPYSYVSDGRTKVSPTGTLTIEEVQ 384
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 361 IYAHESMDIVFECEVTGKPTPTVKKVWNG--DVIIPS-----DNFKIVKEHLQVLGLVK 413
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 385 VDEGAYVCAGMNSAGSSLSKAAL-----KATFETKGRVQKKSKMKG 426
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 414 SDEGFIQCIENDVGNAGAGQLIILEHDVAIPTLPPTSLTSAT--TDHLAPATTGPLPS 471
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 427 QKQKNVQSIIKYLISAVTGNTPAKPP-----PTIEHGHN-- 461
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 472 APRDVVASLVSTRFIKLTWRTPASDPHGNLTYSVFYTKBGVDRERVENTSQPGEMQVTI 531
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 462 QTLVGVSSAILPCQASGK-----PTPGI-SWLRDGLPIDIT-- 496
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 532 QNLMPATVYIFKVMQNHGSGESSAPLRVETQPEVQLPGPAPNIRATATSPTITWTWE 591
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 497 -----DSRISQHSSTGSLHIADLKPKPTGVTYCTIAKNEDGE 531

Db 592 TPLSGNGEIQNYKLYMEKGTDKQDIDVSSH---SYTINGLKKYTEYSFRVVAINKHGP 648
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 532 STWSASLTVEDHTSNAQFVRMPDPSNFPSSPTQPIIVNVTDE-VELHWNAP-STSGAGP 589
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 649 GV-----STQDVAVRTLSDVPSAAPQNLSEVRNKSIVIHWPSSSTTQNGQ 696
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 590 ITGYIIQY----- 597
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 697 ITGYKIRYKASRKSODVTETLVTGTQLSQLIEGLDRGTEYNFRVAALTVNGTGPATDWLS 756
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 598 ---YSPDLGQTFWNPIDYVASTE----- 617
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 757 AETFESDLDET--RVPEVPSSLHVRPLVTSIVVSWTPPENQNIWVRGYAIGYGIGSPHAQ 814
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 618 -----YRIKGLKPSHSMYFVIRAEKNGIGTSPSSALVTSKPAQVALSDKNK 667
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 815 TIKVDYKQRYITENLDPSSSHVITLKAFFNVGEGIPLYESAV---TRPH-----TDTSE 866
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 668 MDMAIAEKRLT-----SEQLIKL-EEVKINSTAVRLFVKRKL---EELIDG--YIKW 716
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 867 VDLFVINAPYTPVPDPTMPMPVGVQASILSHDTIRITWADNSLPKHQKITDSRYITYVRW 926
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 717 RGPPTN---DNQYVNVTSPTSTENYVNSLMPFTNYEFFVI---PYHSGVHSI--HGAPS 768
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 927 ---KTNIPTANTYKKNANA-TTSLYLVGLKPNLTLYEFSVMVTKGRSSSTWSMTAHA-- 979
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 769 NSMDVLTAEPSPSLPPEDVRI--RMLNLTLRISWKAPKADGINGILKGFQIVIVGQAPN 826
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 980 -----TFELVPTSPPKDVTVVSKEGKPRITIVNQPPSE---ANGKITGY-IIIYSTDVN 1030
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 827 NNRNITTNERAASVTFLHVLVTGMT----YKIRVAARSNGGVGVSHTGSEVIMQDTLEKH 882
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1031 AEIHDWVIEPVVGNRLTHQIQELFLDTPTFYFKIQRNRSKGMG---PMSEAVQFR-TPKAD 1086
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 883 LAAQENESFLYGLINKSHVP-----VIVIVAIL 911
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1087 SSDKMPNDQALGSAGKSRPLDGLSDYKPPMSGSNSPHGSPTSPLDSNMLLVIIVSVGVI 1146
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 912 IIFVVIITAYCYWRNSRNSDGKDRSFIKINDGSVEMASN-----NLW-----DV 955
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1147 TIVVVVIAVFCRTTSHQKKRACKSVNGSHKYKGNCKDVKPDPDLWIHHERLELKP 1206
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 956 AQNPQNPNMYNTAGRTMNNRNGQALYSLTP---NAQDFNNCDDYSGTMHRPGSEHHYHY 1013
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1207 DKSPDPNPVMTD---TPIPRNSQ---DITPVDNSMD-----SNIHQRRNSYRGHE 1250
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 1014 AQLTGGPGNAMSTFYGNQYHDDPSPYATTTLVLSNQPAWLNDKMLRAPAMPTNPVPEP 1073
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1251 SE-----DSMSTLAGRR-----GMRPKMM---MPFDSQPPQ 1279
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 1074 -----PARYADHTAGRRSRSSRASDGRGTL-----NGLHHRTSGSQRSDSPPH 1117
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1280 VISAHPIHSLNPHHHFHSSSLASPARSHYHPSSPWPIGTSMSLDRANSTESYRVTPS 1339
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 1118 TDVSYVQ-----LHSSDGTGSSKERTGERPT-----PNKTLMDFIPPPPS 1158
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1340 TDTMPASSQTCTDHDQDEGATSSSYLASSQEESQSLPTAHVRPSHPLKSFAPV-- 1397
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 1159 NPPPPGGHYVDATRRQLNRGSP-----REDYDSVSDGAFARVDVNARPTSRRNL 1211
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1398 -IPPPGPLYDPAL-----PSTPLSQALEPSTHVSKTASIGTLG-RSRPP----- 1443
| : | : | | | | | : | : | | | | | : | : | | | | |
Qy 1212 GGRPLKGRDDDSQRSSLMDDDGGSSEAD 1241
| : | : | | | | | : | : | | | | | : | : | | | | |
Db 1444 --MPVVVSAPEVQETTRMLEDSSESYEPD 1471
| : | : | | | | | : | : | | | | | : | : | | | | |

RESULT 14

Q9V4J9

ID Q9V4J9 PRELIMINARY; PRT: 2016 AA.

AC Q9V4J9;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG17800 PROTEIN.
GN CG17800.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong W., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003841; AAF59271.1; -.
DR HSSP: P40189; 1BQU.
DR FLYBASE; FBgn0033159; CG17800.
DR INTERPRO: IPR000267; -.
DR INTERPRO: IPR001777; -.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; ig; 10.
DR PROSITE: PS00144; ASN_GLN_ASE_1; UNKNOWN_1.
SQ SEQUENCE 2016 AA; 222109 MW; 64A8DE3BB7BD0AB0 CRC64;

Query Match 9.6%; Score 658.5; DB 5; Length 2016;
Best Local Similarity 26.0%; Pred. No. 7.7e-38;
Matches 229; Conservative 130; Mismatches 369; Indels 153; Gaps 33;

Qy 30 PVIIEHPIDVVSRGSPATLNC--GAKPSTAKITWKDQGPVITNKQVNSHRVLDGTS 87
||| : : : | | | : | : | : | : : :
Db 429 PVIRQAFQETMEPGPSVFLKCVAGNPN-TPEISWELDGKKIANNDRYQVGQVYTVNGDV 487
Qy 88 LFLLVNSGKNGKSDAGAYCYVASNEHGEVKSNEGSLKLAILEDVRVRPTVQALGGE 147
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Db 488 VSYLNITS---VHANDGGLYKCIASKVGVA---EHSALNVYGLPYIRQMEKKAIVAGE 541
Qy 148 MAVLECSPPGFPEPVVSWRDKDELRIQDMPRYTLHSDGNLIIDPVDR-SDSGSYQCV 206
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 542 TLIVTC-PVAGYFIDSIVWERDNALPIN--RKQKVPNGTLIIENVERNSDQATYTCVA 598
Qy 207 NNMVGERVSNPARLSVFEPKPK-----FEQEPKDMTVDVGAALVDFCRV-TGDPQPOITWK 260
| | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 599 KNQEGYSARGSLQVMPVQVLPFSFGES----AADVGDIASANCVPKGDLPLEIRWS 654
Qy 261 RKNEMP-----PVTRAYIADNRGLRIERVQPSDEGEYVYARNPAGTLEASHLRVQAP 315
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 655 LNSAPVNGENGFTLVRLNKRISLLNDSLNAFHGRVYKCIATNPAGTSEYVAELQVNPV 714
Qy 316 PSFQTKPADQSVPAAGGTATFECTLVGQPSPAYFWSK-----EGQ-QDLLFPSYVSADGRT 369
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 715 PRWILEPTDKAFAQGSDAKVECKADGFPKQVTVWKAAGVDTGPEYKDLKKSNDIRVE--- 771
Qy 370 KVSPTGTLTIEEVQRVDEGAYVCAGMNSAGSSLSKAALKATFETKGRVQKKSKMGKQKQ 429
||| : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 772 ---EGTLHVDNIQKTNEGYLCEANINGSGLS----- 801
Qy 430 KNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSAILPCQAGSKPTGISWLRD 489
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 802 ---AVIMISVQA-----PPETEKLRNQTARGEPVQLQCEAKGKPIGLWNMN 848
Qy 490 GLPID-----ITDSRISQHSGLHIALDKKPDTPGVYTCIAKNEDGESTWSASLTVE 541
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 849 NMRLDPKNDNRYTIREILSTGVMSLSIKRTSRSDSALFTCVATNAPGSDASINMIVQ 908
Qy 542 DHTSNAQFVRMPDPSPNPSPTQPIIVNVDTTEVLHWNAPSTSGAGPITGYIIQYSPD 601
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 909 E-----VPEMPTALKVLDKSGRSQVLSWAQ-PYDGNPLDRYIEFKRS- 951
Qy 602 LGQTFWNPIDYVA---STEYRIKGLKPSHSYMFVIRAEKIGTIPSVSSALVTTSKPA 658
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 952 -RASWSEIDRVIVPGHTTEAQVQLSPATTYNNIRIVAE--AIGTSQSSEAVTIT---A 1005
Qy 659 QVALSDKNKMDMAIEKRLTSEQLIKLEEVTINSTAVRLFVKKRLEEL---IDGYIK 715
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1006 EEAPSGK-----PQNIKVEPV---NQTMRVTVKPPRTEWNGEILGYVYG 1048
Qy 716 WRGPPRTNDNQYVNTSPSTE-----NYVSNLMFPNTYEFFVPIPHGVSHIHGAP-SN 769
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1049 YK-LSNTSSYVFETINFITEGKEHNLQNLRVYTYQSVVI---QAFNKIGAGPLSE 1103
Qy 770 SMDVLTAEAPSPDPEDVRIMLNLTLRISWAKPKADGINGILKGFQIVIVGQAPNNR 829
||| ||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1104 EEKQFTAEGTSPQSPSDTACTTLTSQTIRVGVWSPPLESANGVIKTYKVV---YAPSD 1160
Qy 830 NITT---NERAASVTLFHLVTGMT-YKIRVAARSNGGVG 865
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1161 YDETKRHYKKTASSDTVLHGLKLYNTYMQVLATAGGDGV 1201

RESULT 15

Q9NBA1

ID Q9NBA1 PRELIMINARY; PRT: 2016 AA.
AC Q9NBA1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DSCAM PRECURSOR.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Schmucker D., Clemens J.C., Shu H., Worby C.A., Xiao J., Muda M.,
RA Dixon J.E., Zipursky S.L.;
RT "Drosophila Dscam is an Axon Guidance Receptor Exhibiting

Search completed: January 22, 2001, 12:53:03
Job time: 1984 sec

```

Qy 30 PVIIIEHPIDVVVSRGSPATLNC---GAKPSTAKITWYKDQGPVITNKQVNSHRIVLDTGS 87
Db 429 PVIRQAQFQETMEPGPSVFLKCVAGGNP-TPEISWELDGKKIANDRYQVQGQVYTVNGDV 487

Qy 88 LFLLVKNSGKNKGDSDAGAYYCVASNEHGEVKSNEGSLKMLREDFRVRPRTVQALGGE 147
Db 488 VSYLNITS---VHANDGGLYKCIASKVGVAA---EHSAKLNVYGLPIRQMEKKAIVAGE 541

Qy 148 MAVLECSPPRGFPEPVVSWRKDDKELRIQDMPRYTLHSDGNLIIDPVDR-SDSGTYQCV 206
Db 542 TLIVTC-PVAGYPIDSIVWERDNRALPIN--RKQKVFPNGTLIENVERNSDQATYTCVA 598

Qy 207 NNMVGERSVNPARLSVFEKPK----FEQEPKDMTVDVGAALVDFCRV-TGDPQPOITWK 260
Db 599 KNQEGYSARGSELEVQVMVPPQVLPSFGES----AADVGDIASANCVVPGKDLPLEIRWS 654

Qy 261 RKNEMP-----PVTRAYIAKDNRLRIERVQPSDEGEYVCYARNPAGTLEASHLRVQAP 315
Db 655 LNSAPIVNGENGFTLVRLNKRSTLLNIDSLNAFHGRGVYKCIATNPAGTSEYVAELQVNP 714

Qy 316 PSFQTKPADQSVPAAGTATFECTLVGQSPAYFWSK-----EGQ-QDLLFPSVYADGRT 369
Db 715 PRWILEPTDKAPAQGSADKVECKADGFPKQPVTKKAVGDTPEGYKDLKSDNIRVE--- 771

Qy 370 KVSPTGTLTIEEVQVDEGAYVCAGMNSAGGSSLSKAALKATFETKGRVQKKSKMGKQKQ 429
Db 772 ---EGTLHVDVQKNTNEGYILCEAINGISGLS----- 801

Qy 430 KNVQSIKYLISAVTGNTPAKPPPTIEHGHQNTLMVGSSAILPCQASGKPTPGISWLRD 489
Db 802 ---AVIMISVQA-----PPEFTEKLRNQTAARRGEPAVLQCEAKGEKPGIGLWNMN 848

Qy 490 GLPID-----ITDSRISQHSSTGSLHIALKPKDPTGVITCIAKNEDGESTWSASLTVE 541
Db 849 NMRLDPKNDNRYYTIREELLSTGVMSLSIKRTERSDSALETCVATNAFGSDDASINMIVQ 908

Qy 542 DHTSNAQFVRMPDPSPNFPSSPTQPIIVNVTDTVELHWNAPSTSGAGPITGYIIQYSPD 601
Db 909 E-----VPEMPYALKVLDRSGRSVQLSWAQP-YDGN SPLDRIIEFKRS- 951

Qy 602 LGQTFWNPIDYVA---STEYRIKLPKPSHSYFVIRAENEKGIGTPSVSSALVITTSKPAA 658
Db 952 -RASWSEIDRVIVPGHTTEAQVKLSPATTYNIRIVAEN--AIGTSQSSEAVTIIT---A 1005

Qy 659 QVALSDKNKMDMAIAEKKRLTSEQLIKLEEVKTIINSTAVRLFVKKRKLLEL---IDGYIYK 715
Db 1006 EEAPSGK-----PONIKVEPV---NQTMRVWTWKPPRTEWNGEILGYVG 1048

Qy 716 WRGPPRTNDNQYVNTVSPSTE-----NYVVSNLMPFTNYEFFVPIYHSGVSHIGAP-SN 769
Db 1049 YK-LSNTNSSYVFETINFITEEGKEHNLEQLNRYVTQYSVVI---QAFNKIGAGLPSE 1103

Qy 770 SMDVLTAEAPSLPPEDVRIRMLNLTTLRISWKAPKADGINGILKGFQIVIVQAPNNNR 829
Db 1104 EEKQFTEAGTSPQPSDTACTILTSQIRVQWVSPPLESANGVIKTYKV---YAPSD 1160

Qy 830 NITT---NERAASVTLFHLVTGMT-YKIRVAASRNSGVGV 865
Db 1161 YDETKRHYKKTASSDVLHGLKKYNTYTMQVLAITAGGDGV 1201

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:18:45 ; Search time 233.01 Seconds
(without alignments)
242.281 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKWKHPFLVMISLLSLSPN.....VLGGYERGEDNNEELEETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8724	100.0	1651	20	Y13566	Human Robo 1 polyp
2	8704	99.8	1649	20	Y08404	Human ROBO1 protei
3	3736.5	42.8	753	20	W83927	Human T85 protein.
4	1592	18.2	1395	20	Y13563	Drosophila Robo 1
5	1592	18.2	1395	20	Y08401	Drosophila sp. ROB
6	1500.5	17.2	1297	20	Y13565	C. elegans Robo po
7	1500.5	17.2	1297	20	Y08403	C. elegans ROBO pr
8	1498.5	17.2	1380	20	Y08402	Drosophila sp. ROB
9	1498	17.2	1381	20	Y13564	Drosophila Robo 2
10	913	10.5	434	20	Y13567	Human Robo 2 polyp
11	913	10.5	434	20	Y08405	Human partial ROBO
12	856	9.8	985	20	Y41716	Human PRO860 prote

13	761	8.7	1257	20	W74152	Human L1 cell adhe
14	708.5	8.1	1571	19	W42087	Human Down syndrom
15	707.5	8.1	1910	19	W42086	Human Down syndrom
16	702.5	8.1	1192	19	W57900	Protein of clone C
17	700.5	8.0	1299	21	Y40439	Human Nr-CAM prote
18	691	7.9	148	20	Y13568	Mouse Robo 1 polyp
19	691	7.9	148	20	Y08406	Mouse partial ROBO
20	682	7.8	1728	12	R13144	Deleted in Colorec
21	681	7.8	1304	19	W59994	Human neural cell
22	667.5	7.7	1447	16	R68553	Deleted in colorec
23	667.5	7.7	1447	20	Y33498	Human DCC protein.
24	664.5	7.6	1028	19	W29667	Homo sapiens DL185
25	644.5	7.4	3117	21	Y53667	Sequence gi/332818
26	596.5	6.8	1018	17	R87028	Human contactin.
27	595.5	6.8	1018	15	R63759	Human contactin (E
28	581.5	6.7	1911	16	R71726	Human PTP-OB. Hom
29	581.5	6.7	1911	18	W27225	Human protein tyro
30	581.5	6.7	1911	20	W94027	Human protein tyro
31	570	6.5	1018	18	W06485	Rat contactin liga
32	569.5	6.5	4412	21	Y53666	Sequence gi/101742
33	568	6.5	1242	19	W52287	Rattus norvegicus
34	556	6.4	1496	20	W81030	Melanoma associate
35	556	6.4	1496	21	Y70469	Human p53 target m
36	551.5	6.3	1225	19	W52289	Homo sapiens cdo t
37	549	6.3	1897	21	Y81785	Human protein tyro
38	549	6.3	1897	21	Y56100	LAR tyrosine phosph
39	521.5	6.0	1139	19	W37779	Rattus norvegicus
40	515.5	5.9	1125	19	W52288	Rattus norvegicus
41	507.5	5.8	1251	19	W37778	Rattus norvegicus
42	506	5.8	1501	16	R72858	Rat receptor type-
43	505	5.8	1070	18	W08747	Human colon carcin
44	478.5	5.5	848	21	Y88565	Human NCAM 140kd i
45	474	5.4	2387	21	Y53665	Mechanical stress

ALIGNMENTS

RESULT 1
Y13566
ID Y13566 standard; Protein; 1651 AA.
XX
AC Y13566;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 1651 AA;

Query Match 100.0%; Score 8724; DB 20; Length 1651;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKWKHPFLVMISLLSPNHLFLAQLIPDPEDVERGNHGTPIPTSDNDNSLGYTGSR 60
Db 1 mkwkhpflvmisllspnhflaqlipdpedvergnhgtptiptsdndnslgytgsr 60
61 LRQEDFPPIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWYKGERVETDKDDPRSHRM 120
Db 61 lrqedfpprivehpsdlivskgepatlnckaegrptptiewykgervetdkddprshrm 120
121 LLPSGSLFFLRIVHGRKSRPDEGVYCVARNYLGEAVSHNASLEVAILLRDDFRQNSDVM 180
Db 121 llpsgslfflrivhgrksrpddegvyccvarnylgeavshnaslevailrddfrqnsdvm 180
181 VAVGEPAVMECPQPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYC 240
Db 181 vavgepavmecpprghpeptiswkkdgsplddkderitirggklmitytrksdagkyvc 240
241 VGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Db 241 vgtnmvgeresevaeltvlerpsfvkrpsnlavtvddsaefkceargdpvptvrwrkddg 300
301 ELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTVQEPHFFVVKPRDQ 360
Db 301 elpkstryeirdhtlkirkvtagdmgsytcvaenmvgkaesatltvqepfhffvvkprdq 360
361 VVALGRVTTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGLTITNVQR 420
Db 361 vvalgrvttfqceatgnppaifwrregsqnlfsyqppqsssrfsvsqtdltnvqr 420
421 SDVGYIYICQLNVAGSIITKAYLEVTDIADIRPPPVIRQGVNQTVAVDGTFLVSCVATG 480
Db 421 sdvgyyicqlnvagsiitkaylevtdiadirpppvirqgvnqtvavdgtflvscvatg 480
481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGTGRYTCIASTPSGEATWSAYI 540
Db 481 spvptilwrkdgvlvstqdsrikqlengvlqiryaklgtgrytciastpsgeatwsayi 540
541 EVQEFQVVPQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSYIEAFS 600
Db 541 evgefqvvpqpprptdnlipsapskpevtvsrntvtlwsqpnlnsgatptsyieafs 600
601 HASGSSWQTVAENVKTETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTQDVLPT 660
Db 601 hasgsswqtvaenvktetsaikglkpnailflvraanaygisdpsqisdpvktqdvlp 660
661 SQGVHDHKVQRELGNVHLHNPVTLSSSSIEVHVTVDQSSQYIQGYKILYRPSGANHGE 720
Db 661 sqgvdhkqvrelgnavlhlhnpvtlssssievhvtvdqssqyiqgykilyrpsganhge 720
721 SDWLVEFVRTPAKNSVVPIDLRKGVNYEIKARPPFNEFGQADSEIKFAKLEEPSAPPO 780
Db 721 sdwlvefvrtpaknsvvpidlrkgvnyeiakarppfnefgqadseikfakleepsappo 780
781 GVTVSKNDGNGTAILVSWQPPEDTQNGMVQYKWCGLGNTRYHINKTVDGSTFSVVP 840
Db 781 gvtvskndgngtailvswqppedtqngmvqykwclgnetryhinktvdgstfsvvp 840

Qy 841 FLVPGIRYSVEVAASTGAGSGVKSEPFQIQLDAHGNPVPSPEDQVSLAQQISDVVKQPAFI 900
Db 841 flvpgirysvevaastgagsgvksepfqqlahgnpvpdpedqvslaqgisdvvkqpafi 900
901 AGIGAACWILMVFSIWLYRHRKRKNGLTSTYAGIRKVPSTFTPTVTYQRGEAVSSGG 960
Db 901 agigaacwilmvfsiwlyrhrkrngltstyagirkvpsftftptvtyrgeavssgg 960
961 RPGLLNISEPAAPWLADTWPNTGNNHDCSISCCTAGNGNSDNLITYSRPADCIANYN 1020
Db 961 rpgllnisepaapwladtwpntgnnhdcsiscctagngnsdnlitysrpadcianyn 1020
1021 NQLDNKQTNLMLESTVYGDVLSKNINEMKTFNSPNLKGFRVNPSPGQPTPYATTQLIQ 1080
Db 1021 nqldnkqtnlmlestvygdvlskninemktnspnlkgfrvnpsgqptpyattqliq 1080
1081 SNLSNNMNGSGDSEKHWKPLGQKQEVAPVQYNIQNKLNKDYRANDTVPTPIPNQ 1140
Db 1081 snlsnnmngsgdsekhwkplgqqkqevapvqyniqnklndyrandtvpttipynq 1140
1141 SYDQNTGGSYNSSDRSGSTSGSGHKGKARTKPVKPGQGMNWADLLPPPAHPHPSNSE 1200
Db 1141 sydqntggssynssdrsgstsgsgkhkgartkpvpkgqgmwadllpppahpphpsnse 1200
1201 EYNISVDESIDQEMPCPVPPARMYLQDELEEEEDERGTPPVVGAASSPAVSYSHQST 1260
Db 1201 eynisvdesydqempcpvpparmylqgdeleeeedergtppvrgaasspaavsyshqst 1260
1261 ATLTPSPQEEELQMLQDCPEETGHMHQHPDRRRQPVSPPPPPIISPHTYGYISGLVS 1320
Db 1261 atltpspqeeelqmlqdcpeetghmhqhpdrrrqpvspppppriisphtygyisglvs 1320
1321 DMDTDAPEEEEDMEVAKMQRLLRLGLEQTPASSVGDLESSVTGSMINGWSASEE 1380
Db 1321 dmdtdapeeeedmevakmqrrllrlgleqtpassvgdlessvtgsmingwsasee 1380
1381 DNISGRSSVSDDSGFFTDADFAQAVAAAAYAGLKVARQMQDAAGRRHFHASQCPRP 1440
Db 1381 dnissgrssvsddsgfftdadfaqavaaaayaglkvarmqmdaagrrhfhasqcprp 1440
1441 TSPVSTDSNMSAAMVQKTRPAKKLKHQPGHLRRETYTDDLPPPPVPPPAIKSPTAQSKTQ 1500
Db 1441 tspvstdsnmsaamvqktrpakkklkhqpghlrrretytdlppppvpppaiksptaqsktq 1500
1501 LEVRPVVVKPLPSMDARTDRSSDRKSSYKGREVLDRGVVDMRTNPGDPREAEQNDG 1560
Db 1501 levrvvvvkplpsmdartdrssdrkssykgrevldrgrvvdmrtnpgdpreaeqndg 1560
1561 KGRGNKAARKDLPPAKTHLIQEDILPYCRPTFTSNPRDPSSSSMSRSGSQRREQA 1620
Db 1561 kgrgnkaarkdlppakthliqedilpycrptftsnprdpssssmsrsgsqrreqa 1620
1621 NVGRRNAEMQVLGGYERGEDNNEELEETES 1651
Db 1621 nvgrrnaemqvlgyyergednneeleetes 1651

RESULT 2

Y08404

ID Y08404 standard; Protein; 1649 AA.

XX

AC Y08404;

XX

DT 24-JUL-1999 (first entry)

XX

DE Human ROBO1 protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
cell morphology; screening assay.

XX

OS Homo sapiens.

XX

PN W09920764-A1.

XX
PD 29-APR-1999.
XX
PF 10-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X08404.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
Claim 1; Page 65-71; 80pp; English.
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1649 AA;

Query Match 99.8%; Score 8704; DB 20; Length 1649;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1649; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 MKWKHVPFLVMISLLSLPNHLFLAQLIPDPEDVERGNDHGTPIPTSDNDDNSLGYTGSR 60
Db 1 mkwkhvpflvmisllslspnhlflaqlipdpedvergndhgtptiptsdndnslgytgsr 60

Qy 61 LRQEDFPPRIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWYKGERVETDKDDPSHRM 120
Db 61 lrqedfpprivehpsdliivskgepatlnckaegrptptiewykgervetdkddprshrm 120

Qy 121 LLPSGSLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVALRDDRQNPSPDM 180
Db 121 llpsgslfflrivhgrksrpdegvyvccvarnylgeavshnaslevalrddfrqnpssdm 180

Qy 181 VAVGEPVMECQPPRGHPEPTISWKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Db 181 vavgepavmecpprghpeptiswkdgsplddkderitirgglkmlitytrksdagkyvc 240

Qy 241 VGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Db 241 vgtnmvgeresevaeltvlerpsfvkrpsnlavtvddsaefkceargdpvptvrwrkddg 300

Qy 301 ELPKSRYEIRDDHTLKIRKVTAGDMGYSYTCVAENMVGAESAATLTVQEPHVFVKPRDQ 360
Db 301 elpkstryeirddhtlkirkvtagdmgsytcaenmvgaesaatlvtqepghvfvkprdq 360

Qy 361 VVALGRTVTTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGDLTITNVQR 420
Db 361 vvalgrtvttfqceatgnppaifwrregsqnlfsyqppqsssrfsvsqtgdltitnvqr 420

Qy 421 SDVGYIICQTLNVAGSIITKAYLEVTDVIADRPVPPVIRQGPVNQTVAVDGTFFVLSVATG 480
Db 421 sdvgyiicqtlnvagsiitkaylevtdviadrpppvirqgpvnqtvavdgtffvlscvatg 480

Qy 481 SPVPTILWRKDGVLVSTQDSRIKLENGVLQIRYAKLGDGTGRYTCIASTPSGEATWSAYI 540
Db 481 spvptilwrkdglvstqdsriklengvlqiryaklgtgrytciastpsgeatwsayi 540

Qy 541 EVQEGFVPVQPPRPTDPNLIPSAKPEVTDVSRNTVTLWQPNLNSGATPTSIIIEAFS 600
Db 541 evqegfvpvpprptdpnlipsakpevtdvsrntvtlswqpnlnsgatptsyieafs 600

Qy 601 HASGSSWQTVAEENVKTETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTDVLPT 660
Db 601 hasgsswqtvaenvktetsaikglkpnailflvraanaygisdpsqisdvpktdvlp 660

Qy 661 SQGVHDHKVQRELGNVHLHNPVTLSSSSIEVHWTVDDQSSQYIQGYKILYRPSGANHGE 720
Db 661 sqgvdhkqvqrelgnavlhlhnpvtlssssievhwtdvqqsqyiqgykilyrpsganhge 720

Qy 721 SDWLVEFVRTPAKNSVVPDLRKGVNYEIKARPFNEFQADSEIKFAKTLLEAPSAPPQ 780
Db 721 sdwlvefvrtpaknsvvipdlrkgvnyeikarpfnefqadseikfaktleapsappq 780

Qy 781 GVTYSKNDGNGTAILVSWQPPPEDTQNGMVQYKVMCLGNTRYHINKTVDGSTFSVVIP 840
Db 781 gvtyskndngntailvswqpppedtqngmvqykvmlcgntryhinktdvgstfsvvip 840

Qy 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQFIQLDAHGNPVSPEQVSLAQGISDVKQPAFI 900
Db 841 flvpgirysvevaastgagsgvkseppqfiqldahgnpvspeqvslagqisdvvpafi 900

Qy 901 AGIGAACWILMVFSIWLRYHRKRRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Db 901 agigaacwilmvfsiwlryhrkrrngltstyagirkvpsftftptvtyrqggeavssgg 960

Qy 961 RPYLLNISEPAAQPLADTWPNPTGNHNDSCISCTAGNNSDNLTTYSRPADCIANYN 1020
Db 961 rpyllnisepaaqpladtwpnptgnhndsciscctagnnsdnlttysrpadcianyn 1020

Qy 1021 NQLDNKQTNLMMLPESTVYGDVDSLKNINEMKTFNSPNLKDGRFVNPSPQPTPYATQLIQ 1080
Db 1021 nqldnkqtnlmmlpestvygdvdslnkinemktnspnlkdgrfvnpspqptpyatt--iq 1078

Qy 1081 SNLSNNMNGSGDSEKHWKPLGQKQEVAPVQYINIVEQNLKNDYRANDTVPTPIPNQ 1140
Db 1079 snlsnnmngsgdsekhwkplgqkqevapvqyniveqnlkndyrandtvptpiynq 1138

Qy 1141 SYDQNTGGSYNSSDRGSSSTSGSGHKGARTPKVPKGGMNWADLLPPPAHPPHNSSE 1200
Db 1139 sydntggsynssdrsgsstsgsgghkgartpkvpkggmwadllpppahpphnsse 1198

Qy 1201 EYNISVDESVDQEMPCVPPARMYLQDDEEEEDERGPTPVGAASSPAVSYSHQST 1260
Db 1199 eynisvdesydeqempcvpparmylqddeeedergptpvgaasspaavsyshqst 1258

Qy 1261 ATLTPSPQEEELQPMQLDCEETGHMQHQRDRRRQVSPPPPPRIPSPHTYGYISGLVS 1320
Db 1259 atltpspqeelqpmqldceetghmqhqrdrvrqvspppprpisphtygyisglvs 1318

Qy 1321 DMDTDAPEEEDEADMEVAKMQTRRLRLGLEQTTPASSVGDLESSVTGSMINGWSASEE 1380
Db 1319 dmdtdapeeeeadmevakmqtrrrllrgleqtpassvgdlessvtgsmingwsasee 1378

Qy 1381 DNISGRSSVSSSDGSFFTDADFAQAVAAAEYAGLKVARQMADAAGRRHFASQCPRP 1440
Db 1379 dnissgrssvssdgsfftdadfaqavaaaeayaglkvarrqmdaagrrhfascprp 1438

Qy 1441 TSPVSTDSNMSAAMVQKTRPAKLLKHQPHLRRETYTDDLPPPPVPPPAIKSPTAQSKTQ 1500
Db 1439 tspvstdsnmsaamvqktrpakllkhqphlrretytdlppppvpppaiksptagsktq 1498

Qy 1501 LEVRPVVVPKLPMSDARTDRSSDRKSSYKGREVLDRGVVDMRTNPGDPREAQEQNDG 1560
Db 1499 levrvvvvpklpsmdartdrssdrkssykgrevldrgvvdmrtnpgdpreaqeqndg 1558

Qy 1561 KGRGNKAARDLPPAKTHLIQEDILPYCRPTFTPSNNPRDSSSSSSSSSGSRGREQA 1620
Db 1559 kgrgnkaardlppakthliqedilpycrptftpsnnprdpsssssssgsrgreqa 1618

Qy 1621 NVGRNIAEMQVLGGYERGEDNNEELETES 1651
Db 1619 nvgrniaemqvlgygergednneeleetes 1649

RESULT 3
W83927
ID W83927 standard; Protein; 753 AA.
XX
AC W83927;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human T85 protein
KW T85; FHMB-6D4; FMHV-SD4; human; neurological disorder; therapy;
KW diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Protein 21..753
FT /label= Mat_protein
FT Region 525..610
FT /note= "has homology to a fibronectin type III domain"
FT Region 638..727
FT /note= "has homology to a fibronectin type III domain"
FT Region 43..101
FT /note= "has homology to a Ig superfamily domain"
FT Region 145..203
FT /note= "has homology to a Ig superfamily domain"
FT Region 237..298
FT /note= "has homology to a Ig superfamily domain"
FT Region 329..394
FT /note= "has homology to a Ig superfamily domain"
FT Region 433..491
FT /note= "has homology to a Ig superfamily domain"
FT Peptide 247..249
FT /note= "RGD motif"
FT Domain 516..600
FT /note= "cytokine receptor homology N-terminal domain"
XX
PN WO9848051-A2.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07714.
XX
PR 10-OCT-1997; 97US-0062017.
PR 18-APR-1997; 97US-0044746.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI Holtzman D, McCarthy SA;
XX
DR WPI; 1999-024021/02.
DR N-PSDB; V69278.
XX
PT New isolated human FTHMA-070 and T85 proteins - used to develop
PT products for the diagnosis and therapy of disorders involving
PT cellular processes, e.g. neuronal development.
XX
PS Claim 31; Fig 3; 127pp; English.
XX
CC This is the amino acid sequence of a novel human protein designated
CC T85, and also referred to as FMHB-6D4 and FMHB-SD4. T85 cDNA (see
CC V69278) was identified in a human foetal brain cDNA library using a
CC screen designed to identify genes encoding proteins having a
CC functional signal sequence. T85 nucleic acids and polypeptides of
CC the invention are useful as modulating agents in regulating a
CC variety of cellular processes. They can be used for identifying
CC compounds which bind to or modulate the activity of the polypeptides

CC (claimed). They can also be used in screening assays, detection
CC assays (e.g. chromosomal mapping, tissue typing, forensic biology),
CC predictive medicine (e.g. diagnostic assays, prognostic assays,
CC monitoring clinical trials, and pharmacogenomics), and methods of
CC treatment (e.g. therapeutic and prophylactic) e.g. for neurological
CC disorders.
XX
SQ Sequence 753 AA;

Query Match 42.8%; Score 3736.5; DB 20; Length 753;
Best Local Similarity 99.4%; Pred. No. 1.6e-188;
Matches 716; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 57 TGSRLRQEDFPPIRIVEHPSDLIVSKGEPATLNCKAEGRPPTIEWYKGGERVETDKDDPR 116
Db 18 sgsrlrqedfpprivehpsdlivskgepatlnckagrptptiewykggervetdkddpr 77

Qy 117 SHRMLLPSSGLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAILRDDRQNP 176
Db 78 shrmlpsgsifflrivhgrksrpdgyvvcvarnylgeavshnaslevailrddfrqp 137

Qy 177 SDVMVAVGEPAVMECQPPRGHPPTISWKKDGSPLDDKDERITIRGKLMITYTRKSDAG 236
Db 138 sdvmvavgepavmecqpprgheptiswkkdgsplddkderitirgklmitytrksdag 197

Qy 237 KYVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWR 296
Db 198 kyvcvgtnmvgeresevaeltvlerpsfvkrpsnlavtvddsaefkceargdpvptvrwr 257

Qy 297 KDDGELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTIVQ---EPPHF 353
Db 258 kddgelpksryeirddhtlkirkvtagdmgsytcvaenmvgkaesatltvqgsepphf 317

Qy 354 VVKPRDQVVALGRTVTFTQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGDL 413
Db 318 vvkprdqvvalgrtvtftqceatgnpqpaiifwrregsqnlfsyqppqsssrfsvsqtdl 377

Qy 414 TITNVQRSDVGYIICQTLNVAGSIITKAYLEVTVDIADPPPIRQGPVNQTVAVDGTfv 473
Db 378 titnvqrsdvgyiicqtlnvagsiitkaylevtvdiadppppirgqgvnqtvavdgtfv 437

Qy 474 LSCVATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDTGRYTCIASTPSGE 533
Db 438 lscvatgspvptilwrkdglvstqdsrikqlengvlqiryaklgdtgrytciastpse 497

Qy 534 ATWSAYIEVQEFQVGPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTS 593
Db 498 atwsayievqefgvvpvqpprptdnlipsapskpevtvdsrntvtlwsqpnlnsgatpts 557

Qy 594 YIIAFAFSGSSWQTVAENVKTETSAIKGLKPNALFLVRAANAYGISDPSQISDPVK 653
Db 558 yileafshasgsswqtvaenvktetsaikglkpnailflvraanaygisdpqisdvpk 617

Qy 654 TQDVLPTSGVDHKQVQRELGNVLAHLNPTVLSSSSIEVHWTVDQSQYIQGYKILYRP 713
Db 618 tqdvlptsgvdhkqvqrelgnavlhlhnpvtlssssievhwtvdqssqyiggykilyrp 677

Qy 714 SGANHGESDVLVFEVRTPAKNSVVPDLRKGVNYEIKARPPFNEFGADSEIKFAKTLEE 773
Db 678 sganhgesdvlvfevrtpaknsvvpdlrkgvnyeikarpffnefgadseikfaktlee 737

RESULT 4
Y13563
ID Y13563 standard; Protein; 1395 AA.
XX
AC Y13563;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 1 polypeptide.
XX


```

RESULT 7
Y08403
ID Y08403 standard; Protein; 1297 AA.
XX
AC Y08403;
XX
DT 24-JUL-1999 (first entry)
XX
DE C. elegans ROBO protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;

```

KW cell morphology; screening assay.

XX

OS Caenorhabditis elegans.

XX

PN W09920764-A1.

XX

PD 29-APR-1999.

XX

PF 20-OCT-1998; 98WO-US22164.

XX

PR 14-NOV-1997; 97US-0971172.

XX

PR 20-OCT-1997; 97US-0062921.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Goodman CS, Kidd T, Mitchell KJ, Tear G;

XX

DR WPI; 1999-312615/26.

XX

DR N-PSDB; X57252.

XX

PT Robo polypeptides, a new immunoglobulin superfamily member

XX

S Claim 1; Page 59-63; 80pp; English.

XX

CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from *Drosophila* sp.,
CC *C. elegans*, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.

XX

SQ Sequence 1297 AA;

Query Match 17.2%; Score 1500.5; DB 20; Length 1297;
Best Local Similarity 32.1%; Pred. No. 7e-71;
Matches 408; Conservative 166; Mismatches 480; Indels 219; Gaps 36;

Qy 68 PRIVEHPDLSVSGEPATLNCKAERPTPTIEWYKGGERVETDKDDPRSHMLPSGSL 127

Db 30 pviiehpdivvsvrgspatlnckag-pstakitwykdgqpvitnkeqvnshriyldtgs 88

Qy 128 FFLRIVHGRKSR-PDEGVYVCVARNYLGEAVSHNASLEVAIRDDFRQNPDSVMVAVGEP 186

Db 89 filkvnsqngkgsdagaycvasnehgevsneksiklamlredfrvrptvqalgem 148

Qy 187 AVMECPQPRGHPEPTISWKKDGSPLDDK-ERITRG-KGLMITYTRKSDAGKYVCVGTN 244

Db 149 avlecsprgrfpepvswrkdldkrligdmprytlhdsnglildpvdrsdsgtyqcvann 208

Qy 245 MVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWKDDGELPK 304

Db 209 mvgervsnparlsvfekpkfeqepkdmtdvgaavlfdrvtgdpqgqitwkrknepmpv 268

Qy 305 SR-YEIRDHTLTKIRKVTAGDMGSYTCVAENMVGKAEASATLTVEPPEHVKPRDQVVA 363

Db 269 trayiakdnrglriervqpsdegeyvcyarnpagtleasahlrvqappsftkpadqsvp 328

Qy 364 LGRTVTFCQCATGNPQPAIFWRREGSQNLFF-SYQPPQSSSRFSVSQTGLDITNVQRSD 422

Db 329 aggtatfectlvqpspayfwskeggqdlfpsy--vsadgrtkvsptgtltieevrqud 386

Qy 423 VGYICQTLNVAGSIITKAYLE-----VTDV 448

Db 387 egayvcagmsnagslskaalkatfetkgrvgkkskmgkqkgnvgsiikylisavtgn 446

Qy 449 IADRPPIVIRQGPVNTVAVDGTFVLSCVATGSPVPTILWRKDGVLVSTQDSRIKQLENG 508

Db 447 tpakppptiehgqngtlmvgssailpcqasgkptpslswldgldpidtdsrisqhshtg 506

Qy 509 VLIQIRYAKLGDGTGRYTCIASTPSGEATWSAYIEVQEGFVQPPRPPTDPNLIPSAPSKE 568

Db 507 shliadlkkpdtgvtyciaknedgestwsasltvedhtsnaqfvrmppdpnfpsptapi 566

Qy 569 VTDVSRNTVTLWSQPNLNSGATP-TSYIIIEAFSHASGSSWQTVANVKTETSAIKGLKPN 627

Db 567 invvtdtevelhwnapstgagpitygilyqyspdlggtwfnipdyvasteyrikglkps 626

Qy 628 AIYLFVRAANAYGISDPQISDPVKT-----QDVLPTSQGVHDHVKQVREL-GNAVHLHL 681

Db 627 hsymfviraenekgigtspvssalvttskpaqvalsdknmdmaiaekrltseqlikle 686

Qy 682 NPTVLSSSSIEVHWTVDQSQYIQYKILYR-PSGANHGESDMLVEFVRTPAKNSVVIDP 740

Db 687 evktinstavrlfwkkrkleelidgyykwgpprtdnq---yvntspstenyvvsn 742

Qy 741 LRKGVNYEIKARPF---FNEPQAGSEIKFAKLEEAPSAPPQVTGSKNDGNGTAILVS 797

Db 743 lmpftnyeffvipyhsgvhsihgapsnmdvltaeappslppedvirmli--nlrtiris 800

Qy 798 WQPPPEDTQNGMVQYKVVCLNETRYHINKTVDSGTFVSVIPPLVPGIRYSVEVAAGT 857

Db 801 wkakpadingilkqfivivgqapnnrnttneraasvtfhlvtgmytkirvaarsn 860

Qy 858 AGSGVKSEPFQIQLDAHGNVSPEDQVSLAQQISDVVKQPAFIAGIAGAACWI-----I 910

Db 861 ggvgv-----shgtsevinnqdtlekhlaaqgenesflyglinkshvpvivai 910

Qy 911 LMVFSIWLYRHRKRKRLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGGRPGLNLISE- 969

Db 911 liifviiiaicywnsrnsd---gkdrsf-----ikindgsvhmasn--nlwdvaqn 958

Qy 970 PAAQPLADTWPTGNHNDSCISCTAGNNSDNLTTYS---RPADCIANYNQLDN 1025

Db 959 pqnqpmntagrmtnnrgqalysltpnadqffnccdygqtmhrpgsehhyhaqltg 1018

Qy 1026 KQTNLMPESTVIGDVLNKNINEMKTFNSPNLKDGRFVNPSGQPTPYATTQLIQSNLSN 1085

Db 1019 gpgnam---stfyg-----ngyhd-----dpspyatttlvlsn--- 1048

Qy 1086 NMNNGSGDSGRKHWKPLGQQKQEVAPVQYINIVEQNLNKDYRANDTVPTIPYNQSYDQN 1145

Db 1049 -----qq-----pawln---dkmlrapamptnvpvpepp--aryadh 1080

Qy 1146 TGGSYNSDGRSS-----TSGSQ-----GHKK-----G 1168

Db 1081 tagrrsrssrasdgrgtlmgllhrtsgsqrdspphtdvsyqlhssdgtgsskertge 1140

Qy 1169 ARTPKVPKQGGMNWADLLPPPPAHPPPHSNSEYINISVDESIDQEMPCVPPARMYLOQD 1228

Db 1141 rrtpt--pnktlm---dfippppsnppp-----pgghvy---d 1169

Qy 1229 ELEEEEDERGPTP 1241

Db 1170 tatrqlnrgstp 1182

RESULT 8

Y08402

ID Y08402 standard; Protein; 1380 AA.

XX

AC Y08402;

XX

DT 24-JUL-1999 (first entry)

XX

DE *Drosophila* sp. ROBO2 extracellular domain protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX

OS *Drosophila* sp.

XX

PN W09920764-A1.

XX

PD 29-APR-1999.

XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57251.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 52-56; 80pp; English.
XX
SC This invention describes novel Robo (roundabout) polypeptides, involved
in nerve guidance which have been isolated from *Drosophila* sp.,
C. elegans, human and murine samples. The products of the invention can
be used to raise anti-Robo antibodies, which can be used to modulate cell
function or morphology. The Robo polynucleotides and fragments are useful
as probes and primers and for production of the Robo polypeptides. The
probes and primers are also useful in screening assays.
XX
SQ Sequence 1380 AA;

Query Match 17.2%; Score 1498.5; DB 20; Length 1380;
Best Local Similarity 29.0%; Pred. No. 9.6e-71;
Matches 410; Conservative 221; Mismatches 526; Indels 259; Gaps 42;

Qy 68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTPTIEWYKGGERVETDKDDPRSHRMLLPSGSL 127
Db 4 priiehpmdttvpkndpftfncgaegnpptptiqwfkdgrelktdtg---shrimlpaggl 60

Qy 128 FFLRIVHGRKSRPDEGVYVVCARNYLGEAVSHNASLEVAAILRDFRQNPDMVMVAVGEPA 187
Db 61 fflkvihsrr-esdagtywceaknefgvarsnatlqvavlrdefrlepantrvaggeva 119

Qy 188 VMEQPPRGPHEPTISWKKDGSPLD---DKDERITIRGGKMLITVTRKSDAGKYVCVGTN 244
Db 120 lmeagprgspepqiswrkngqtnlvgnkiriri-vdggnlaiqearqsdgryqcvvkn 178

Qy 245 MVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRK---DDGEL 302
Db 179 vvgtresataflkvhrpfllrgpqnqtavvgssvfvqriggdplpdlvrrtasggm 238

Qy 303 P-----KSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGKAEASATLVQEPHP 352
Db 239 plrkfswlhasgrvvhvledrslkldvtdlmedgeytceadnavggitatgiltvhappk 298

Qy 353 FVVKPRDQVVALGRVTFTQCEATGNPQPAIFWRREGSQNLFSYQPPQSSRSFSVQTGD 412
Db 299 fvirkpnlqveigdevlfecqanghrprtlywsvegnslll---pgyrdgrmevltpe 355

Qy 413 ----LTIITNVQRSDVGYII-CQTLNVAGSIITKAYLEVTDIADRPPIVIRQGPVNQTV 467
Db 356 grsvlsiarfaresdgvvtcnalnavgssvrtvsv-dtqfelpplieggpvnqtlp 414

Qy 468 VDGTFVLSCVATGSPVPTILWRKDGVLVTSQDSRIKQLEN-GVLQIR-YAKLGDTRYTC 525
Db 415 vksivlpcrtlgtppvqswyldgipidqveherrnlsdagaltisdigrhedeglytc 474

Qy 526 IASTPSGEATWSAYIEVEFGVPVPPRPTDNL-----IPSAPSKEPTVDSRN 575
Db 475 vasnrngksswsgylrld-----tptnplnkffrapelstypgpgkpmvkegen 525

Qy 576 TVTSLW-OPNLNGATPTSIIIEAFSHASGSWQTVAVENVTETSAIKGLKPNAILFLV 634
Db 526 svtlswtrsnkvsgsslvgyviemfgknetdgwvavgrvqntttftgtllpgvnyfili 585

Qy 635 RAANAYGISDPSQISDPVKTDVLPTSQGVHDHKVQRE-LGNAVLHLNPTVLSSSSIEV 693

Db 586 raenshgslslpspmsepi-tvgtryfngsdlsearallsgdvelsnasvdstsmkl 644

Qy 694 HWTVDQSQYIQGYKILYR----- 712

Db 645 twqi-ingkyvegfyvyarqlpnpivnnpavtsntnpllgststsasasasalistk 703

Qy 713 ----PSGANHGSEDWLVEFVRTP-----AKNSVVPDLRKGVNYEIKARPF 754

Db 704 pnlaaagkrdgetnqsgggaptlntkyrmltilngggassctitglvqytlveffivpf 763

Qy 755 FNEFQADSEIKFAKLEAPSAAPPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQEYK 814

Db 764 ykveqgkpsnsriartledvpseapygmeal11--nssavflkwkapelkdrhgvllnyh 821

Qy 815 VWCLGNTRYHI-----NKTVDGSTFSVVPFVLPGIRYSVEVAASTGAGSGVKSEQFI 869

Db 822 vlvrgidtahnsriltntvtdaasptlvlanitegvmytvgaagnnagvgpycpvatl 881

Qy 870 QLDHAGNPVSPEDQVSLAQISDVVQPAFIAGIGAACWILWFSIWLY---RHRKRN 926

Db 882 rldpitrldp--finqrhvdvltqpwfiillgallavmlsfamfvkrkhmmmq 936

Qy 927 GLTSTYAG----IRKVPSTFTPTPTVYQGGAEV---SSGG---RPLGLNISEPAAQ 974

Db 940 salntmrnghtsdvklmpsls-----arnngywdssdggmvrpspggdslemkd 992

Qy 975 WLADTWPNNGNNHNDSCISCTAGNDSNLTYSRPADCIANYNNQLDNQKTNLMLE 1034

Db 993 hiadypvcgagpgaggtssggsggagsg---asggdihghgserngqrry---- 1044

Qy 1035 STVYGDV-DLSNKNEMKTFNSPNLKGFRVNPSPQPTPYATTQLIQSLNLSNMNN--- 1089

Db 1045 ----geysniptdyaevsfgkapseygrhgnas--papyatssilshqggqggqpryq 1098

Qy 1090 ----GSGDSSEKHKKPLGQQKQEVAPVQYVQYQNLKNDYRANDTVPPTIPYQNS--- 1141

Db 1099 qrpvpyglqipmh--phyqqqqh---qqqagqthqhqalqghqqlppsnlyqgmstt 1153

Qy 1142 ---YDQNTGGS-----YNSSD-----RGSSSGSQG--- 1164

Db 1154 seiypntgprsvyseqyyypkdkqrhihienklsnchtyeaapgakqsspsissqfasv 1213

Qy 1165 -----HKKGARTPKVPKQGGMNWADLL-----PPPPAHP 1193

Db 1214 rrqqlppncsigresarfkvintdqknqgnlldidgssmcyngladsgcgsspspmaml 1273

Qy 1194 PPHNSSEYINISVDESYDQEMPCVPVPPARMYLQDELEEEEDERGPTPPVGAASSPAV 1253

Db 1274 mshedeahlytadgldd-----merlykvkvdqppqqqqqlplvpqhaeghlq 1326

Qy 1254 SYSHQSTATLTPSQEELQPMQLDCPEETHGMQHQP 1289

Db 1327 swrngstrssrkngge-----cikepseliyap 1354

RESULT 9

Y13564

ID Y13564 standard; Protein; 1381 AA.

XX

AC Y13564;

XX

DT 30-JUL-1999 (first entry)

XX

DE *Drosophila* Robo-2 polypeptide.

XX

KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
modulation; nerve cell function.

XX

OS *Drosophila* sp.

XX

PN W09925833-A1.

XX

```

RESULT 10
Y13567
ID Y13567 standard; Protein; 434 AA.
XX
AC Y13567;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.

```



```

Qy   360 QVVALGRVTVFQCETAGNPPQAFLFWRREGSQNLFFSYQPQQSSSRFSVSQTGDLTITNVQ  419  
      |:| |::|| ||:|| :|:||||| | | :| | | | |||||:|  
Db    1 qivaggrvtvfpcetkgnppafwfkqegsqnllfpncqqgpnsrcsvsptgdltitniq  60  
  
Qy   420 RSDVGYYICQLTNWAGSIITRAYLEVDVIADRPPPVIROGPVNQTVAVDGTFTVLSCVAT  479  
      ||| ||||| | ||||: || |||||: ||||:| || | |:||| :| || |  
Db    61 rsdagyyicqlatvagsilakaqltvdldtrpppiilggpanqtlavdgallckcat  120  
  
Qy   480 GSPVPPTILWRKDGVLVSTQDSRIKLQENGLVLIQIRYAKLBDTRYTCIASTPSGEATWSAY  539  
      | |:| | | |:| :| | | ||: :| ::| | |:|:~ ~| | :| |:  
Db    121 gdplpvswlkegf tprdrpratiqeqgtlgiklnrlidstgytcvatssgeaswsav  180  
  
Qy   540 IEVQEFGVPVQPPRPDTNP LIPSAPSKPEVTDVSRNTVTLSWQNPNLNGATPTSIIIEAF  599  
      ::| :| | :| :| :| | | | | | | | | | | | | | | | | | | |  
Db    181 ldvtsgatis--kn ydlsdlpgpps kqvdtvkns vtlswqp gtpglpasayieaf  238  
  
Qy   600 SHASGSSWQTVAENVKTETSAIKGLKPNAIYLFLVRANAYGISDPQSISDPVKTKDVLP  659  
      | :| :| | | :| | :| :| | | | | | | | | | | | | | | | |  
Db    239 sqvswnsqwtvanhvktlytvrglrntplyfmvrain-----  277  
  
Qy   660 TSQGVDHKVQRQRELGN AVLHLHNPTVLSSSSI EVHWTVDDQSOYIQGY KILYPGSANGH  719  
  
Db    278 -----  277  
  
Qy   720 ESDWLVEFVRTPAKN SVVIPDRKG VNYEIKARPFFNEFOGADSEIKFAKTL EEAPSAPP  779  
  
Db    278 -----pk  279
```

Qy	780	QGVTVSKNDGNGTAILVSWQPPPPEDTQNGMVEYKVCWLGNTRYHINKTVDGSTFSVVI	839
Db	280	vsvtqxx-----	286
Qy	840	PFLVPGIRYSVEVAASTGAGSGVKSEPFQIQLDAHGNVPSPEDQVSLAQQISDVVKQPAF	899
Db	287	-----pq-----	288
Qy	900	IAGIGAACWIIILMVFSIWLYRHRKRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSG	959
Db	289	-----knng-----	292
Qy	960	GRPGLLNISEPAAQPLADTWPNTGNHNDCSISCTAGNGSNDNLTTYSRPADCIANY	1019
Db	293	-----stwan-----	297
Qy	1020	NNQLDNKQTNLMLPESTVYGDVLSNKNINEMKTFNSPNLKDGRFVNPFSGQPTYATTQLI	1079
Db	298	-----	297
Qy	1080	QSNLSNNMNGSGDSGEKHWKPLGQQKQEVAPVQYNIVEQKNLKNKYRANDTVPTTIPYN	1139
Db	298	-----	297
Qy	1140	QSYDQNTGGSNSSDRGSSTSGSQGHKGARTPKVPKQGGMNWADLLPPPAHPPHSNS	1199
Db	298	-----vp-----lppppvqplpgtel	313
Qy	1200	EEYNISVDES-YDQEMCPVPPARMYLQ---DELEEEEDERGPTPPVRGAASSPAAVSY	1255
Db	314	ehyaveqgengydsdswcpplpvqtylhqgledel-eeddrvptppvrgvassp-aisf	371
Qy	1256	SHQSTATLTPSQEELQFMLQDCPEETGHMQHPDRRRQPVSPPPPPRIPSPHTYGYIS	1315
Db	372	ggqstatltpspreemqpmqasp-----	395
Qy	1316	GPLVSDMDTDAPEEEEDADMEVAKMQTRRLRLRGLEQTPASSVGDLESSVTGSMINGWG	1375
Db	396	-----	395
Qy	1376	SASEEDNISSGRSSVSSSDGSFFTDADFAQAVAAAAEYAGLKVARRQMADAAGRRHFHAS	1435
Db	396	-----xftss	400
Qy	1436	QCPRPTSPVSTDSNMSAAVMQKTRPAKKLKHPG	1469
Db	401	qrprptspfstsdntsaalsqacprptkhhkq	434

```

RESULT 11
Y08405
ID Y08405 standard; Protein; 434 AA.
XX
AC Y08405;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human partial ROBO2 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens.
XX
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.

```

Qy 1020 NNQLDNKQTNLMLPESTVYGDVDLSNKINEMKTFNSPNLKDGRFVNPSGOPTPYATTOLI 107

Db 401 qrprptspfstdsntsaalsqsrprptkkhkqg 434

PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.


```

RESULT 13
W74152
ID W74152 standard; Protein; 1257 AA.
XX
AC W74152;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human L1 cell adhesion molecule.
XX
KW Human L1 cell adhesion molecule; I
KW nervous system development; nerve
KW neuronal cell cohesive interaction
XX

```

```

OS Homo sapiens.
XX
PN US5872225-A.
XX
PD 16-FEB-1999.
XX
PF 18-NOV-1994; 94US-0341843.
XX
PR 26-JUN-1992; 92US-0904991.
PR 18-NOV-1994; 94US-0341843.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Lemmon V;
XX
DR WPI; 1999-166719/14.
DR N-PSDB; X01598.
XX
PT Human L1 cell adhesion molecule - supports neurite outgrowth and is
PT involved in nervous system development and repair
XX
PS Claim 1; Fig 3; 45pp; English.
XX
CC This sequence is the human L1 cell adhesion molecule (L1CAM) of the
CC invention. L1CAM supports growth of neurites in vitro and is involved in
CC development of the human nervous system and in nerve regeneration. It is
CC useful in in vivo and in vitro experiments on nerve growth and
CC regeneration. L1CAM mediates cohesive interactions of neuronal cells to
CC each other and to extracellular matrix.
XX
SQ Sequence 1257.AA;

Query Match 8.7%; Score 761; DB 20; Length 1257;
Best Local Similarity 24.9%; Pred. No. 4.3e-32;
Matches 283; Conservative 159; Mismatches 422; Indels 274; Gaps 45;

Qy 14 LLSLSPNHLFLAQLIPDPEDVERGNHDGTP IPTSDNDNSLGYTSRLRQEDFP PRIVEH 73
Db 11 lllcsp-----clllqipeeyeghvhmepvit-----eqsprllvfvf 46

Qy 74 PSDLIVSKGEPATLNCKAEGRPPTPIETWYKGGERVETDKD-----DPRSHRMLPSGS 126
Db 49 ptdddi-----slkceasgkpevqfrwtrdgvhfkpkeelgvtyvqshpgsfittgnn 101

Qy 127 LFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAAILRDDFRPNP SD----VMVA 182
Db 102 snf-----agrfqgilyrcfasnklgtamsh----eirImaegapkwketvkvpeve 149

Qy 183 VGEPAVMEQPPRGHPPEPTISWKDGSPLDDKDERITI-RGGKLIMITYRKSD-AGKYVC 240
Db 150 egessvvlpcnpppsaeplriywmnskilhikqderivtmggngnlyfanvltsdnhsdyic 209

Qy 241 ----VGTMVGERESEVAELTVLPRSPFVKRPSNLAVTVDDSAE-----FKCEAR 286
Db 210 hahfpgrttiigkep--idlrvkatsmidrkrllfptnssshlvalggqplvleciae 267

Qy 287 GDPVPTVRWRKDDGELPKSRYEIRD-DHILKIRKVTAGDMSGYTCAENMVKGAEASATL 345
Db 268 gfptptikwlrlpsgpmprdvrtvynhntklqlklvgeeddgeyrcslaenslgarhayyv 327

Qy 346 TVQEPHFVVKPRDQVVALGRTVTVFQCEATGNPQPAIFWRREGSQNLLFSYQPPQSSSRF 405
Db 328 tveaapywihkpsqshlypggetarldcqvgqrpqpewtwing-----ipeelakdqky 382

Qy 406 SVSQTGDLTIITNVQSDVGYIYCQTLNVAGSIITKAYLEVTVDIADRP PPVIRQGPVNQT 465
Db 383 ri-qrgalilsnvqpsdtmvtqcearnrhglllanaylyvvql----pakiltad-nqt 435

Qy 466 -VAVDG-TFVLSCVATGSPVPTILW-RKDGVLVTSQDSRIKQLENGVLQIRYAKLGD TGR 522
Db 436 vmavagastavllckafapvpsvowldedattv-laderffpvanatqlairldandtar 494

```

```

Qy      64 EDFPPRIVEHPSDLIVSKGEPATLNCKAEGRPTPTIEWYKGGERVETDKDDP-----RSRH 119
      || ::| :| :| || :| :| ||| | | | | | | | | | | | | | | | | | |
Db      403 edgtpkiisafsekvsvpaepwslmcnkvgtlptitw-----tlddpilkggshr 454

Qy      120 ---MLLPSGSLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAI LRDDFRQNP 176
      | :| ::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      455 isqmitsegnvlsylnissqvr-dggyvrcannasg-vlyqarinv---rgpasirp 509

Qy      177 SDVMVAV-GEPAVMCEQPPRGHPEPTISWKKDGSPLDDKDERITI-RGGKLMITYTRKS- 233
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      510 mknitaiaagrđtyihr-vigypysikwyksnllpfnhrqvafenngtlklisdvqkev 568

Qy      234 DAGKYVCVGTNMVGERE---SEVAELTVLERPSFVKRPSLAVTVDDSAEFC- EARGDP 289
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      569 degeytc---nvlvqpqlstsqsvhvtv-kvppfiqpfefrfsigqrvfipcvvvsqdl 624

Qy      290 VPTVRWRKDDGELPKSRYEIRDD---HTLKIRKVTAGDMGSYTCVAENMVGKAESATL 345
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      625 pititwqkdgrppsglgtidnidftsslrissnlsmhngnytcianeeaaavehqsq 684

Qy      346 TVQEPHPPFVVKPRDOVVALGRVTVFQCEATGNPQPAIFWRREGSQNLFS-----YQP 398
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      685 ivrvppkfvvqprddgdyigkavilncsaegyvpvtlvwk-----fsgagvypqfq 736

Qy      399 PQSSSRFSVSQTGDLTITNVQSDVGYYICQTLNVAGSSIITKA-YLEVTVDIADRP PPVI 457
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      737 ialnigrigvlsngsllikhvveedsgyyklcvsnvdgadvksmyltv-----kipami 790

Qy      458 RQGPVNQTVAVDG-TFVLSCVATGSPVPTILWRKDGVLVSTQDSR--IKQLENG-----V 509
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      791 tsyp-ntlatggqkkmestahgeklpiivrwekedriinpemarylvstkevgivist 849

Qy      510 LQIRYAKLGDIGRYTCIASTPSGEATWSAYIEVQEFGVPVQPPRPDPNLPISAPSKPEV 569
      || | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      850 lqilptvredsgffschainsygedrgliqltvqe-----ppdppelei 893

Qy      570 TDVSRNTVTLSWQPNLNSGATPTSIIIEAFSHASGSSWQTVAENVK-----TETSAIKGL 624
      || | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      894 kdvkartitlrwtmqfmcnsplitavdieckn--ksdswds-aartkdvspolnsatiidi 950

```

RESULT 14
W42087
ID W42087 standard; Protein; 1571 AA.
XX
AC W42087;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human Down syndrome-cell adhesion
XX
KW DS-CAM2; Down syndrome-cell adhesion
KW signal transduction; trisomy 21; m
KW holoprosencephaly; corpus callosum
KW schizencephaly; diagnosis; assay;
XX
OS Homo sapiens.
XX
PN W09817795-A1.
XX
PD 30-APR-1998.
XX
PF 23-OCT-1997; 97WO-US19547.
XX
PR 25-OCT-1996; 96US-0029322.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Korenberg JR;
XX
DR WPI; 1998-271791/24.
DR N-PSDB; V31988.
XX

```

RESULT 15
W42086
ID W42086 standard; Protein; 1910 AA.
XX
AC W42086;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human Down syndrome-cell adhesion molecule DS-CAM1.
XX
KW DS-CAM1; Down syndrome-cell adhesion molecule; neural cell;
KW signal transduction; trisomy 21; mental retardation;
KW holoprosencephaly; corpus callosum agenesis;
KW schizencephaly; diagnosis; assay; human.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
Peptide 1..23
FT /label= Sig_peptide
FT Protein 24..1910
FT /label= Mat_protein
FT Domain 24..887
FT /label= IG
FT /note= "immunoglobulin type-C2 domain"
FT Domain 888..1594
FT /label= Fbn
FT /note= "fibronectin type III domain"
FT Domain 1595..1616
FT /label= Transmembrane
FT Domain 1617..1910
FT /label= Cytoplasmic
FT Region 24..126
FT /label= Ig1
FT Region 127..225
FT /label= Ig2
FT Region 226..316
FT /label= Ig3
FT Region 317..409
FT /label= Ig4

```

```

FT      Region      410..506
FT      /label= Ig5
FT      Region      507..603
FT      /label= Ig6
FT      Region      604..697
FT      /label= Ig7
FT      Region      698..792
FT      /label= Ig8
FT      Region      793..887
FT      /label= Ig9
FT      Disulfide-bond 46..102
FT      Disulfide-bond 145..197
FT      Disulfide-bond 246..293
FT      Disulfide-bond 335..385
FT      Disulfide-bond 428..484
FT      Disulfide-bond 525..575
FT      Disulfide-bond 617..669
FT      Disulfide-bond 711..766
FT      Disulfide-bond 809..865
FT      Disulfide-bond 1307..1359
FT      Modified-site 78..80
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 106..108
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 470..472
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 487..489
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 658..660
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 666..668
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 710..712
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 748..750
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 795..797
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 924..926
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1142..1144
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1160..1162
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1250..1252
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1271..1273
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1324..1326
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1341..1343
FT      /note= "Asn is N-glycosylated"
FT      Modified-site 1488..1490
FT      /note= "Asn is N-glycosylated"
XX
PN      WO9817795-A1.
XX
PD      30-APR-1998.
XX
PF      23-OCT-1997; 97WO-US19547.
XX
PR      25-OCT-1996; 96US-0029322.
XX
PA      (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI      Korenberg JR;
XX
DR      WPI; 1998-271791/24.
DR      N-PSDB; V31981.
XX
PT      New isolated Down's Syndrome-cell adhesion molecule - used to
PT      develop products for detection, diagnosis and therapy of
PT      developmental and neurological abnormalities

```

Search completed: January 22, 2001, 12:19:37
Job time: 1734 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:26:10 ; Search time 325.28 Seconds
(without alignments)
344.638 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKWKHPFLVMISLSPN.....VLGGYGERGDNNEELETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	8315	95.3	1651 2 T14160	transmembrane rece
2	8120	93.1	1612 2 T30805	duttl protein - mo
3	2607.5	29.9	1344 2 T14316	rig-1 protein - mo
4	1515	17.4	1273 2 T42405	sax-3 protein - Ca
5	852	9.8	423 2 T29549	hypothetical prote
6	764.5	8.8	1260 1 S05479	neural cell adhesi
7	761	8.7	1257 1 A41060	neural cell adhesi
8	757.5	8.7	1443 2 I50600	neogenin - chicken
9	739.5	8.5	1259 2 S36126	neural cell adhesi
10	726.5	8.3	1427 2 I51669	tumor suppressor -
11	713	8.2	1259 2 A43425	Bravo/Nr-CAM cell
12	707.5	8.1	1896 2 T08851	Down syndrome cell
13	702.5	8.1	1028 2 I58164	BIG-1 protein - ra
14	695.5	8.0	1028 2 A53449	plasmacytoma-assoc
15	693	7.9	1268 1 A39640	neural cell adhesi
16	684.5	7.8	874 2 T29548	hypothetical prote
17	678.5	7.8	2222 2 T13924	sdh protein - frui
18	667.5	7.7	1447 2 A54100	tumor suppressor p
19	653.5	7.5	1040 2 A49356	transient axonal g
20	652	7.5	1040 2 A34695	axonal glycoprotei
21	644.5	7.4	5175 2 T20992	hypothetical prote
22	644.5	7.4	5198 2 T43290	hemocentin precurs
23	637	7.3	1277 2 T30532	neural cell adhesi
24	633.5	7.3	1036 2 S22383	axonin 1 precursor
25	629.5	7.2	1272 2 S26180	neurofascin - chic
26	628	7.2	1239 1 A32579	neuroglian - fruit
27	618.5	7.1	1209 2 T42718	probable neural ce
28	613	7.0	1018 2 JC4211	neural adhesion pr
29	596.5	6.8	1018 2 A54744	contactin 1 precur

30	596.5	6.8	1907 2 S50893	protein-tyrosine-p
31	589.5	6.8	1010 2 JU0094	Fil protein precur
32	589.5	6.8	1091 2 S01998	contactin precurs
33	585	6.7	1232 2 T43027	neural cell adhesi
34	584	6.7	1912 2 A56178	protein-tyrosine-p
35	583	6.7	2029 1 TDFFLK	protein-tyrosine-p
36	581	6.7	1894 2 C54689	protein-tyrosine-p
37	573	6.6	1020 2 S05944	neuronal cell surf
38	571.5	6.6	1021 2 A57112	contactin precurs
39	570	6.5	1256 2 T03096	CDO protein - rat
40	555.5	6.4	1898 2 S46216	leukocyte antigen-
41	551.5	6.3	1240 2 T03097	CDO protein - huma
42	549.5	6.3	7962 2 I38346	elastic titin - hu
43	549	6.3	1897 1 TDHULK	leukocyte antigen-
44	548.5	6.3	1265 1 A37967	neural cell adhesi
45	548	6.3	1197 2 T30581	neural cell adhesi

ALIGNMENTS

RESULT 1
T14160
transmembrane receptor protein Robol - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14160
R:Kidd, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodman, C.
Cell 92, 205-215, 1998
A>Title: Roundabout controls axon crossing of the CNS midline and defines a novel sub
A:Reference number: Z17897; MUID:98117249
A:Accession: T14160
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1651 <KID>
A:Cross-references: EMBL:AF041082; NID:g2811215; PID:g2811216; PIDN:AAC39960.1
C:Function:
A:Description: appears to function as the gatekeeper controlling midline crossing
C:Keywords: transmembrane protein

Query Match 95.3%; Score 8315; DB 2; Length 1651;
Best Local Similarity 94.5%; Pred. No. 0;
Matches 1560; Conservative 42; Mismatches 49; Indels 0; Gaps 0;

Qy	1	MKWKHPFLVMISLSPNHLFLAQLIPDPEDVERGNHGTPTSDNDNSLGYTGSR 60
Db	1	MKWKHLPLLMISLTLKKHLLAQLIPDPEDLERNNGTPTSDNDNSLGYTGSR 60
Qy	61	LRQEDFPPRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHRM 120
Db	61	LRQEDFPPRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHRM 120
Qy	121	LLPSGSLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAIRDDFRQNPDSVM 180
Db	121	LLPSGSLFFLRIVHGRKSRPDEGVYICVARNYLGEAVSHNASLEVAIRDDFRQNPDSVM 180
Qy	181	VAVGEPAYMECQPPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Db	181	VAVGEPAYMECQPPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Qy	241	VGTNMVGERSEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Db	241	VGTNMVGERSEKADVTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Qy	301	ELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAMNMGKAEASATLTQVEPPHFVVKPRDQ 360
Db	301	ELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAMNMGKAEASATLTQVEPPHFVVKPRDQ 360
Qy	361	VVALGRVTTFQCEATGNPQPAIFWRREGSNLLFSYQPPQSSSRFSVSQTGLDITNVQR 420
Db	361	VVALGRVTTFQCEATGNPQPAIFWRREGSNLLFSYQPPQSSSRFSVSQTGLDITNVQR 420

Qy 421 SDVGYICQTLNVAGSIITKAYLEVTDVIADRPVIRQGPVQTVAVDGTFLVSCVATG 480
Db 421 SDVGYICQTLNVAGSIITKAYLEVTDVIADRPVIRQGPVQTVAVDGTFLVSCVATG 480

Qy 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIATSPSGEATWSAYI 540
Db 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIATSPSGEATWSAYI 540

Qy 541 EVQEGFVQVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSIIIEAFS 600
Db 541 EVQEGFVQVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSIIIEAFS 600

Qy 601 HASGSSWQTVVAENVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTQDVLP 660
Db 601 HASGSSWQTVVAENVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTQDVLP 660

Qy 661 SQGVHDKQVQRELGNVHLNHPVTLSSSSIEVHWTVDDQSQYIQGYKILYRPSGANHGE 720
Db 661 TQGVHDKQVQRELGNVHLNHPVTLSSSSIEVHWTVDDQSQYIQGYKILYRPSGANHGE 720

Qy 721 SDWLVEFVRTPAKNSVVPIDLRKGVNEYKARPPFNEFGADSEIKFAKTEEAPSPAPQ 780
Db 721 SEWLVEFVRTPAKNSVVPIDLRKGVNEYKARPPFNEFGADSEIKFAKTEEAPSPAPQ 780

Qy 781 GVTVSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWLGNETRYHINKTVDGSTFVSVP 840
Db 781 SVTVSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWLGNETRYHINKTVDGSTFVSVP 840

Qy 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQIQLDAHGNVPSPEDQVSLAQQISDVVKQPAFI 900
Db 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQIQLDSHGPNVSPEDQVSLAQQISDVVKQPAFI 900

Qy 901 AGIGAACWIIIMVFSIWLRYRHRKRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Db 901 AGIGAACWIIIMVFSIWLRYRHRKRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960

Qy 961 RPLGLNISEPAAQPLWADTWNTGNNHNDSCISCTAGNNSDNLTTYSRPADCIANYN 1020
Db 961 RPLGLNISEPATQPLWADTWNTGNNHNDSCINCTAGNNSDNLTTYSRPADCIANYN 1020

Qy 1021 NQLDNKQTNLMLPESTVYGDVLSNKNINEMKTFNSPNLKDGRFVNPSSGQPTPYATTQLIQ 1080
Db 1021 NQLDNKQTNLMLPESTVYGDVLSNKNINEMKTFNSPNLKDGRFVNPSSGQPTPYATTQLIQ 1080

Qy 1081 SNLSNMMNNGSGDSEKHKWPLGQKQEVAPVQYINVEQNKLNKDYRANDTVPTPIYINQ 1140
Db 1081 ANLINMMNNGSGDSEKHKWPPGQKQEVAPVQYINVEQNKLNKDYRANDTVPTPIYINQ 1140

Qy 1141 SYDQNTGGSYNSSDRSGSTSGSQGHKKGARTPKPVQGGMMWADLLPPPAHPPPHSNSE 1200
Db 1141 SYDQNTGGSYNSSDRSGSTSGSQGHKKGARTPKPVQGGMMWADLLPPPAHPPPHSNSE 1200

Qy 1201 EYNISVDESVDQEMPCVPPARMYLLQDELEEEEDERGPTPPVGAASSPAVSYSHQST 1260
Db 1201 EYSMSVDESVDQEMPCVPPARMYLLQDELEEEEAERGPTPPVGAASSPAVSYSHQST 1260

Qy 1261 ATLTPSPQEEELQMLQDCPEETGHMHPDRRRQPVSPPPPPRISPSPHTYGYISGPLVS 1320
Db 1261 ATLTPSPQEEELQMLQDCPEELGHMHPDRRRQPVSPPPPPRISPSPHTYGYISGPLVS 1320

Qy 1321 DMDTDAPEEEDEADMEVAKMQTRRLRLRGLQTPASSVGDLESSVTGSMINGWSASEE 1380
Db 1321 DMDTDAPEEEDEADMEVAKMQTRRLRLRGLQTPASSVGDLESSVTGSMINGWSASEE 1380

Qy 1381 DNISGRSSVSSSDGSFFTDADFAQAVAAAAAYAGLKVARRMQDAAGRRHFHASQCPRP 1440
Db 1381 DNISGRSSVSSSDGSFFTDADFAQAVAAAAAYAGLKVARRMQDAAGRRHFHASQCPRP 1440

Qy 1441 TSPVSTDSNMSAAVMQTRPAKKLKHQPGHLRREYTTDDLPPPPVPPPAIKSPTAQSKTQ 1500
Db 1441 TSPVSTDSNMSAAVIQKARPTKQKHQPGHLRREYTTDDLPPPPVPPPAIKSPVSQSKAQ 1500

Qy 1501 LEVRPVVVKPLPSMDARTDRSSDRKGSYKREVLDRQVVDVMDRTNPGDPREAQEQNDG 1560

Db 1501 LEARPTMGPKLASIEARADRSSDRKGSYKREALDGRQVTDLRTSPGDPREAQEQNEG 1560

Qy 1561 KGRGNKAAKRDLPKATHLIQEDILPYCRPTFTPTSNPRDPSSSSSSSRSGSGSRQREQA 1620
Db 1561 KARGTKAKRDLPKATHLIPEDILPYCRPTFTPTSNPRDPSSSSSSSRSGSGSRQREQA 1620

Qy 1621 NVGRRNIAEMQVLGGYERGEDNNEELETES 1651

Db 1621 NVGRRNMAEMQVLGGFERGEDNNEELETES 1651

RESULT 2

T30805

dutt1 protein - mouse

N;Alternate names: transmembrane receptor protein Robol homolog

C;Species: Mus musculus (house mouse)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C;Accession: T30805

R;Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.

submitted to the EMBL Data Library, July 1998

A;Description: The mouse homologue of human DUTT1/H-robol gene: protein sequence and

A;Reference number: Z20879

A;Accession: T30805

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1612 <WU>

A;Cross-references: EMBL:Y17793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1

A;Experimental source: brain

C;Genetics:

A;Gene: dutt1

A;Map position: 16

Query Match 93.1%; Score 8120; DB 2; Length 1612;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1525; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

Qy 57 TGSRLRQEDFPPIRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPR 116
Db 18 SGRSLRQEDFPPIRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPR 77

Qy 117 SHRMLPSGSLFFLRIVHGRKSRPDEGVYCVARNYLGEAVSHNASLEVALRDDFRQNP 176
Db 78 SHRMLPSGSLFFLRIVHGRKSRPDEGVYCVARNYLGEAVSHNASLEVALRDDFRQNP 137

Qy 177 SDVMVAVGEPVMECQPPRGHPPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAG 236
Db 138 SDVMVAVGEPVMECQPPRGHPPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAG 197

Qy 237 KYVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWR 296
Db 198 KYVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWR 257

Qy 297 KDDGELPKSRYEIRDDHTLKRKVTAGDMGSYTCVAENMVGAASATLTVQEPHFVVK 356
Db 258 KDDGELPKSRYEIRDDHTLKRKVTAGDMGSYTCVAENMVGAASATLTVQEPHFVVK 317

Qy 357 PRDQVVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGDLTIT 416
Db 318 PRDQVVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGDLTIT 377

Qy 417 NVQSDVGYICQTLNVAGSIITKAYLEVTDVIADRPVIRQGPVQTVAVDGTFLVSC 476
Db 378 NVQSDVGYICQTLNVAGSIITKAYLEVTDVIADRPVIRQGPVQTVAVDGTFLVSC 437

Qy 477 VATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIATSPSGEATW 536
Db 438 VATGSPAPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIATSPSGEATW 497

Qy 537 SAYIEVQEGFVQVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSII 596
Db 498 SAYIEVQEGFVQVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSII 557

```

Qy 58 GSRLRQEDFPFPRIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRS 177
   |||: | | |||| | | ||||| ||||| | | | | | :|||:
Db 32 GSRVGPEDAMPRIVEQPPDLVVSERGEATLPCRAEGRPRPNIEWYKNGARVATAREDPRA 91

Qy 118 HRMLPSSGSLFFLRIVHGGRKSAPDEGVYVCVARNYLGEAVSHNASLEVAILLRDDFRQNPS 177
   ||:|||||:| | ||||| ||||| ||||| | | | | | :|||:| | |||||:|
Db 92 HRLLLPSSGALFFPRIVHGRRSRPDEGVYTCVARNYLGAASARNASLEVAILLRDDFRQSPG 151

Qy 178 DVMVAVGEPAMVECPQPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGK 237
   :|:|||||||:| | ||||| :||| | :||:|||||||:| | ||||
Db 152 NVVVAVGEPAMVECPVPPKGHPVLVTKKGKILKEEGRITIRGGKLMMSHTFKSDAGM 211

Qy 238 YVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVWRWK 297
   |:|: | | |||| | | |||||:| | | | | |:| | | | | |:|
Db 212 YMCVASNMAGERESGAELVVLERPFLRRPINOVLADAPVNLCEVQGDQPQNLHWK 271

Qy 298 DDGELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGAESAATLTVQEPHFVVKP 357
   ||||| | |||| |:| | :||: | | |||||:| | |||| |:| | |
Db 272 DDGELPAGRYEIRSDHSLWIDQVSSDEGTYTCVAENSVGRAESAHSVHPQFVTKP 331

Qy 358 RDQVVALGRTVTFQCEATGNPQPAIFWRREGSONLLFSYQPPQSSSRFSVSQTDGLTITN 417
   :|: | | | | | | |||||:| | | | | | | | | | | | | |
Db 332 QNQTVAPGANVSFQCEQKGNPPPAIFWQKEGSQVLLFPSQSLQPMGRLLVSPRGQLNITE 391

Qy 418 VQRSDVGYYICQTLNAGVSIITKAYLEVTDIVADRPPVIRQGVNQTVAVDGTFLVLCV 477
   |: | | | | | | :|||:| | | | | | | | | | | | | |
Db 392 VKIGDGGYYVCQAVSVAGSILAKALLEIKGASIDGLPPIILQGPANQTLVLGSSVWLPCR 451

Qy 478 ATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDTRYTCIASTPSGEATWS 537
   |:| | | | | | | | | | | | | | | | | | | | | | | |
Db 452 VIGNPQPNIQKKDERWLQGDSDQFNLMNDGTLHTASIQEMDGFYSVCARSSIGEATWN 511

Qy 538 AYIEVQE-FGVVPQPPRPDTPNLIPSAPSKPEVTVSRNTVTLWSQPNLNSGATPTSYYI 596
   :||: | | | | | | | | | | | | | | | | | | | | | |
Db 512 SWLRKQEDWG--ASPGPATGSPNPPGPSQPIVTEVTANSITLWKPNNPQSGATATSYVI 569

Qy 597 EAFSHASGSSWQTVANVKTETSIAIKLKPNAIYFLVRAANAYGISDPSQISDPVKTDQ 656
   |||| |:|:|:|:|:|: | | | | | | ||||| |:|:|:|:|:|:|:|:|:|:|
Db 570 EAFSQAAGNTWRTVADGVQLETTYISGLQNPITYLFLVRAVGAWGLSEPPSVPEVQTDQ 629

Qy 657 VLPTSQGVDRKQVQRELGNVHLHNPVLTVSSSSIEVHWTVDOQSQYIQGYKILYRPSGA 716
   | | | | | | | | | | | | | | | | | | | | | | | |
Db 630 SLSRPAEDPWKQRLGAEEAVRMQEPPTVLGPRTLQVSWTVLQGVQVQGFVRWSRTAGL 689

Qy 717 NHGSDWLVPFEVRTPAKNSVVPDLRKGVNYEIKARPFNEFQADSEIKFAKTLEEAPS 776
   : | | | | | | :|||:| | | | | | | | | | | | | |
Db 690 DQG--SWTMDLQLQSPHKQSTVLRLGLPPGAQIQIKVQVQOQEGLGAESEPFVTRSIEPEAPS 747

Qy 777 APPQGVTVSKNDGNGTAILVSWQPPPEDTQNGMVQYKIVWCLGNETRYHINKTVDGSTFS 836
   |||| | | | | | | :|||:| | | | | | | | | | | | | |
Db 748 GPPOGVAVALGDRNSSVTVSWEPPLPSORNGVITEYOIICLGNESRFLNRSAGAWGS 807

```



```
RESULT 6
S05479
neural cell adhesion molecule L1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
```

```

Qy      56 YTGSRRLQEDFPFPRIVEH-PSDLIVSKGEPATLNCKAEGRPPTPTIEWYKGGVERVETDK-- 112
      || : : || | : | : : | : || | : :
Db      26 YKGHHVLE---PPVITEQSPRRLVVFPTDDISLKCEARGRPQVEFRWTKDGIHFKPKEEL 82

Qy     113 ----DDPSRHRMLLPSGSLFFLRIVHGRKSRPDGEGVYVCVARNYLGEAVSHNASLEVAI 167
      : | : : : : : : : : : : : : : : : : :
Db      83 GVVVHEAPYSGSFTIEGNSFAQRF-----QGIYRCYASNKLGTAMSH---EIQL 129

Qy     168 LRDDFRQNPDS----VMVAVGEPAMVECPGRGHEPTISWKKDGSPLDD--KDERITI- 220
      : : : : : | | | : | : | | | | : : | : | : :
Db     130 VABGAPKWPKEITVPKPEVEEGESVVLPCNPPPSAAPPRIYWM--NSKIFDIKQDERVSMG 187

Qy     221 RGGKLMITYTRKSD-AGKYVCVGTNMVGER---ESEVAELTVLERSPVKRPSNLAVTVD 276
      : | : | : | : | : | : | : | : | : | : | :
Db     188 QNGDLYFANVLTSDNHSDYIC-NAHPGTRITIQKEPIDLVRKPTNSMDRKPRLLEPTN 246

Qy     277 DSAE-----FKCEARGDPVPTVRWRKDDGELPKSRYEIRDDH--TLKIRKVTAGD 324
      : : : : : | : | : | : | : | : | : | : | :
Db     247 SSSRLVALQQCSLILCEIAGEGFPTPTIKWLHPDMPMTDRV-IYQNHNTQLQNLVGEED 305

Qy     325 MGSYTCVAENMYGKAEASATLTVQEPHPFVVKPRDQVVALGRVTVFQCEATGNPQPAIFW 384
      | | | : | : | : : : : : : : : : | : | : | :
Db     306 DGEYTCIAENSLSGARHAYYVTEAAPYWLKQPSHLYGPGETARLDCQVQGRPQPEITW 365

Qy     385 RREGSQNLLFSYQPPQSSRSFSVSTGDLTITNVQRSDVGYIQCITLNVAGSIITKAYLE 444
      : : : : : : : : : : : : : : : | : : | : :
Db     366 RING-----MSMETVNKDQKYRIEQ-GSLILSNVQPTDLMVTQCEARNQHGILLANAYY 419

Qy     445 WTDVIADRPFPVIRQGPVNOT-VAVDG-TFVLSCVATGSPVPTLLWRKDGVLVSTQDSRI 502
      : : : : : | : | : | : | : | : | : | : | :
Db     420 VVOL-----PARILLKD--NOTYMAVEGSTAYLLCKAFGAPVPSVONLDEBGTTLQDEREF 473

```

```

RESULT 7
A41060
neural cell adhesion molecule L1 precursor - human
N;Alternate names: L1CAM
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: A41060; S18454; A35331; S21971; S21972; A60223; A31072; G02506
R;Hlavin, M.L.; Lemmon, V.
Genomics 11, 416-423, 1991
A;Title: Molecular structure and functional testing of human L1CAM: an interspecies comp
A;Reference number: A41060; MUID:92120663
A;Accession: A41060
A;Molecule type: mRNA
A;Residues: 1-1257 <HLA>
A;Cross-references: GB:M64296; NID:q186053; PIDN:AAC14352.1; PID:q3068548

```

Query Match 8.7%; Score 761; DB 1; Length 1257;
Best Local Similarity 24.9%; Pred. No. 2.6e-28;
Matches 283; Conservative 159; Mismatches 422; Indels 274; Gaps 45;

Qy 14 LLSLSPNHLFLAQLIPDPEDVERGNDHGTP IPTSDNDDNSLGYTGSR LRQEDFP PRIVEH 73
|| || | || | : | : | : | | | | |

```
Qy      57 TGSRLRQDFFPP-RIVEHPSDLIVSKGEPATLNCKAEGRPRTPIIEWYKGGERVETDKDDP 115  
    ||| : | . | : | : | : | : | : | : | : |  
Db      9 TGSVVR---TFTPFYLVEPMDDLVSRGASVIMNCSSYCETPKIEWKKGDTLLNLNVSDD- 65  
  
Qy     116 RSHRMILLPGSGLFLFRIVHGKRSRPDDEGVYVCVAR-NYLGEAVSHNASLEVAILLRDDFRQ 174  
    | ||| ||| : | : : |||| | ||| | | | | | | |  
Db     66 --RRQLLPDGSLLINSVVHSKHKNPDEGYQCVAVESLGSISVRTAKTLVAGL-PRFTS 122  
  
Qy     175 NPSDVMVAVAGEPAVMCEQPGRGHPEPTIISWKKDGPLDDKDERITIRGGKLMIYTRKSD 234  
    | | |' | : : | : | : | : | | | | | : | : |  
Db     123 QPELSSVIKGN SAILNCE-VNVLDAPFVRWEQDRQPLSLDDRVLKPLSGALLIGNATDTD 181  
  
Qy     235 AGKYVCVGVTNMVGERESEVAELTVLERPS-----FVKRPSNLAVTVDDSAEFKCEARGD 288  
    | | | : : | : ||| : | | : ||: ||| : | : | | |  
Db     182 GGFYRCVIESGTPKYSEEAECLKLPDPPEEQSQLVFVRQPSLSIKVTQONAVFPVCVAGGF 241  
  
Qy     289 PVPTRWRKDDGEL---PKSRYEIRDHTLKIRKVTAGDMGSTCYCAENMVGKAESAATL 345  
    | | ||| : : | | : : | : | | : |: ||| : | : | | |  
Db     242 PTPIYRWTKNGEELITEDSERFALRAGGSLILSDVTEEDVGTYTCIADNETIEAQABL 301  
  
Qy     346 TVQEPPHFVVKPRDQVVALGRVTTFQCEATGNPQAIFWRREGSQNLFSYQPPQSSSRF 405  
    || | | : : : : : | : || | | | : : | : | : |  
Db     302 AVQVPPEFLKRPANITAHEMSDIVFECEVTGKPTPTVKWKNGDVVIPSDY-----F 353  
  
Qy     406 SVSQTGDLTIINVORSVDGYYICQLNLVAGSIITKAYLEVTDVIADRP--PPVIRQGPVN 463  
    : : : : : : || : | | | : | : | : | : | | |  
Db     354 KIVKEHNQLVGLVKSDEGFYQCIAENDVGNAQAQAQLIILDVLAIPTLPSTSLTSATN 413  
  
Qy     464 QTVAVDGTFLVSCVATGSPVPTILWRKDGLVYSTQDSRIKQLENGVLQIRYAKLGDTGRY 523  
    : | | | | | | | | | | | | | | : : | : |  
Db     414 DHLA-----PARTGPLPTAPRDVVATLVST-----RFTRL----- 443  
  
Qy     524 TCIASTPSGSEATWSAYIEVQEGFVQVQPPRPTDNPILSPAPSKEPTDVDSRNTVLSWQP 583  
    : | | | | | | | | | | | | | | : : | : : | : |  
Db     444 -----TW R-----TPVSDPQ--GDNLITYSIFYTKE--GINRREVENTSRP 479  
  
Qy     584 NLNSGATPTSIIIEAFSHASGSSWQTVAENVKTSETSA-IGKLKPNAIYLFVLRANAAYGI 642  
    | | | | | | | | | | | | | | | | : | : | : | : |  
Db     480 G-----ETQVMIQMLMPETVYVFRVVAQNKHGH 507  
  
Qy     643 SDPSQSISDPVKYTQDVLPTSGVDHKQVQRELGNVHLHNLPTVLSSSSIEVHW-TVDQQS 701  
    : | : | : | : | : | : | : | : | : | : | : | : |  
Db     508 GESSA---PLX----VATQPEV---QLPGPAPNIRAYAGSPT-----SVTVTWETPLSGN 552  
  
Qy     702 QYIQGYKILYRPSGANHGESDWLVEFYRTPAKNSVVIPLDRKGVNWEIKARPFNEFGQA 761  
    : : | : | : | : | : | : | : | : | : | : | : | : |  
Db     553 GEIQNYKLIYMKGQD-SEQDQDV-----AGLSYTTIGLKKYTESFRVYVAYNKHGPV 605  
  
Qy     762 DSEIKFAKLTLEAPSAPPQGVTVSKNDGNGTAILVSWQPPPEDTQNGMVOEYKVMCLGNE 821  
    : : | : | : | : | : | : | : | : | : | : | : | : |
```

```
Qy      1 MKWKHVFFVLIMSLLSLSPNHLFLAQLIPDPEDVERGNDHGTPIPTSDNDDNSLGYTGRS 60
       || | : | : | : | || |
Db      4 MLWYVFLPLLSCPLLIQ-----IPDE-----YKGHH 30

Qy     61 LRQEDFPFRIVEH-PSDLIVSKGEPAILNCKAEGRPPTTIEWYKGGSERVETDK----- 112
       : : | | : | : | : | | | | | : :
Db    31 VLE---PPVITEQSPRRLLVFPTDDISLKCEAGRGPQVEFRWTKDGIHFKEELGVVVH 87

Qy    113 DDPRSHRMLLPSGLSFFLRIVHGKRSRPDEGVYVCVARNYLGEAVSHNASLEVAILRDDF 172
       : | : | : | : | : | : | | | | | : : :
Db    88 EAPYSGSFTIEGNNSFAQR-----QGIIKYASNNLTGMASH---EIQLVAEGA 134

Qy   173 RQNPSD----VMVAVGEPAYMECQPPRGHPEPTISWKDGSPLDKDERITI-RGGKLMI 227
       : : | : | | | | | | | | | | : | | : : :
Db   135 KPWKETVKPVEVEEGESVVLPCNPPTSAAPLRIYWMSKILHIKQDERVSMGONGDYLF 194

Qy   228 TYTRKSD-AGFYVCVTNMVGER---ESEVAELTVLERPSFVKRPNSLAVTVDDSAE--- 280
       || | : | : | : | : | : | : | : | : | :
Db   195 ANVLTSDNHS DYIC-NAHPGTRTTIQKEPIDLRVKPTNSMIDRKPRLLFPTNSSSHLVA 253

Qy   281 -----FKCEARGDPVPTVRWRKDDELPKRSRYEIRDDH--TLKIRKVTAGDMGYSTCV 331
       : | : | : | : | : | : | : | : | : | : | :
Db   254 LQGQSLILBCIAEGFPTPTIKWLHPSDPMPDRV-IYNHNKTQLLNVGEEDDGEGYTCL 312

Qy   332 AENMVGAKEASATLTVQEPHPFVVKPRDQVVALGRVTTFQCEATGNPQPAIFWRREGSQN 391
       || | : | : | : | : | : | : | : | : | : |
Db   313 AENSLGSARHYIYVTVEAAPYWLQKPQSHLYGPGETARLDCQVQGRQPQEVWTRING--- 369

Qy   392 LLFSYQQPQSRSRFVSQTGDLITINQRSVDGYICYCTLNVAGSIITKALEYLTVDIAD 451
       | : | : | : | : | : | : | : | : | : | :
Db   370 --MSIEKVNKOQXRIEQ-GSLILSNWPQSDTMVTCQEARNQHGILLANAIYIVVQL--- 423

Qy   452 RPPPVIHQGPVNT-VAVDG-TFVLSCVATGSPVPTILWRKDGVLVTSQDSRIKQLENGV 509
       : : : : | : | : | : | : | : | : | : |
Db   424 -PARILTKD- NOTYMAVEGSTAYLLCKAFAGPVSWQLDEEGTIVLQDERFFPYANGH 480
```


Qy 510 LQIRYAKLGDGTGRYTCIASTPSGEATWSAYIEVQEGVFPVQPPRPT----- 555
| | : | | | | | : | | : | | |
Db 481 LGIRDLQANDTGRYFCQAANDQNNVITLANLQVKEATQITQGRSTIEKKGARVFTTCQA 540
Qy 556 --DPNL----- 559
| | : | | | | | : | | : | | |
Db 541 SFDPSLQASITWRGDRDLQERGDSDKYFIEDGQLVIKSLDSDQGDYSCVASTELDEVE 600
Qy 560 -----IPSAPSKPEVTD--VSRNTVTLWQPNLNSGATPTSIIIE-AFHASHGSS 606
| | : | | | | | : | | : | | |
Db 601 SRAQLLVGSGPGVPVPHLELSDRHLLKQSQVHLSWSPAEDHNSPIEKYDIEFEDKEMAPEK 660
Qy 607 WQTVANVKETSAIKGLKPNAYLFLVRAANAYGISDPSQISDPVKTQDVLPTSQGVHD 666
| | : | | | | | : | | : | | |
Db 661 WFLSGKVPNGQSTTLKLSPTVHYTFRVTAINKYGPGEPSVSETVVTEAAPEKNPVDV 720
Qy 667 KQVQRELGNVHLHNPVTLSSSSIEVHWTVQSQYIQGYKILYRPSGANHGESDWLVF 726
| | : | | | | | : | | : | | |
721 RGEGETNNMVI-----TWKPLRW-MDWNAPQIQ-YRQWRPLGK---QETW--- 762
Qy 727 EVRTPAKNSVVPDLRKGVNYEIKARPFNEFOGADSEIKFAKTLLEEAPSAPP--QGTV 784
| | : | | | | | : | | : | | |
Db 763 KEQTVSDPFLVSNSTSFVPEYELKVQAVNNQKGPEQVTIGYSGEDYQVSPLEEDITI 822
Qy 785 SKNDGNGTAILVSWQPPEDTQNGMVQYKVV--WCLGNTRY---HINKT---VDGSTFS 836
| | : | | | | | : | | : | | |
Db 823 F---NSSTVLVRWRPVDLAQVGHRLGYNVITYWKGSRKSKRHVHKSMMVVPANTTS 878
Qy 837 VVIPFLVPGIRYSVEVAASTGAGSGVKSE-----PQFIQLDAHG----- 875
| | : | | | | | : | | : | | |
Db 879 AILSGLRPSYHVEVQAENGRGLGPASEWTFSTPEGVPGHPEALHLEQSDTSLLLHWQ 938
Qy 876 -----NPVSPEDQVSLAQISDVVVKQPAFIAGIAGAACWILMVFSIWLY 919
| | : | | | | | : | | : | | |
Db 939 PPLSHNGVLTGYLLSYHPLDGESEKQLFFNLSD----- 971
Qy 920 RHRKKRNLSTYAGIRKVPSTFT-TPTVTVYQRGGEAV-----SSGGRPGLLNISEPA 971
| | : | | | | | : | | : | | |
Db 972 -PELRTHNLNLNLDLQ---YRFQLQATTHQGEAIVREGGTMAFGKPDGFGNISVTA 1026
Qy 972 AQPWLADTW-PNTG 984
| | : | | | | | : | | : | | |
Db 1027 GENYSVSVWVPREG 1040

RESULT 10

1669

umor suppressor - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: I51669

R:Pierceall, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.
Dev. Biol. 166, 654-665, 1994

A:Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the

A:Reference number: I51668; MUID:95113183

A:Accession: I51669

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1427 <PI>

A:Cross-references: EMBL:U10986; NID:g606873; PIDN:AAA70168.1; PID:g606874

C:Genetics:

A:Gene: XDCCA

Query Match 8.3%; Score 726.5; DB 2; Length 1427;
Best Local Similarity 20.6%; Pred. No. 1.3e-26;
Matches 348; Conservative 201; Mismatches 584; Indels 553; Gaps. 55;

Qy 71 VEHPDSDLVSKGEPATLNCKAEG-RPTTIEWYKGERVETDKDDPSRSHMLLPSGSLFF 129
| | : | | | | | : | | : | | |
Db 43 LSEPSDAVTMRGNNVNLCSAQSDRGAPIIKKKDKGVYLNVLVIDE---RRQLPSGSFLI 99
Qy 130 LRIVHGRKSRPDEGVYVCVAR-NYLGEAVSHNASLEVA---ILRDDFRQNPDSVMVAVG 184

Db 100 QNVVHSRHRPDEGVYQCEASLDSVGTIVSRATKVLVAGPLRLIL-----SQTESVTAFIG 154
Qy 185 EPAVMECOPPRGHPEPTISWKKDGSPLDDKDERIT-----IRGGKLMITYTRKSDAG 236
| | : | | | | | : | | : | | |
Db 155 DTALLRCE-ITGEPMTISWQK-----NEEDLKVTGDPRLVLPSTGLQISRLQATDGG 208
Qy 237 KIVCVGTNNVGERESEVAELTVLERPS-----FVKRPSNLAVTVDDSAEFKCEARGDPV 290
| | : | | | | | : | | : | | |
Db 209 VTRCLAKNPGSARVGNELRLSESGLHRQQVFLQRPNSNVVAIEGQDAVLECAVSGYPT 268
Qy 291 PTVRWKDDGELP--KSRYEIRDHDLTKIRKVTAGDMGSYTCVAENMVGAASATLTQV 348
| | : | | | | | : | | : | | |
Db 269 PTIVWQGDPEVPPIRTKYSVLGGSNLLISNVTDDAGATYCNATYKNENTSFSAIDLTVM 328
Qy 349 EPPHFVVKPRDQVVALGRVTFTQCEATGNPQPAIFWRREGSQNLLFSYQPPSSSRFSVS 408
| | : | | | | | : | | : | | |
Db 329 VPPQFLNHPANLYAYESMDIEFECAVSGKPSPTVKWTNGEVVPSDY-----FQIV 380
Qy 409 QTGDLITNVQSRSDVGYICQTLNVAGSIITKAYLETVDIADRPVPPVIRQGPVNTVAV 468
| | : | | | | | : | | : | | |
Db 381 DGSNLRILGLVKSDEGYQCIAENEAGNIQTYAQLIIPD----- 419
Qy 469 DGTFLVLSCVATGSPVPTILWRKDGVLSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIAS 528
Db 420 ----- 420
Qy 529 TPSGEATWSAYIEVQEGVFPVQPPRPTDNLIPSAPSKPEVTVSRNTVTLWQPNLNSG 588
| | : | | | | | : | | : | | |
Db 420 -----PAVPSLPSAPRDVVPVLSRFRVLSWRPPEVSK 456
Qy 589 ATPSYIIEAFSHAGSSWQTVANVKTETS-----AIKGLKPNAYLFLVRAANAYG 641
| | : | | | | | : | | : | | |
Db 457 GNIQTYTVY-----FSKQGVQRERAVNTSQPISLQITVGNLTPEETYNFRVAYNEWG 509
Qy 642 ISDPSQISDPVK--TQDVLPTSQGVHDHQQVQRELGNVHLHNPVTLSSSSIEVHWTVQ 699
| | : | | | | | : | | : | | |
Db 510 ---PGESSQEVKVTQPELQVPGPVENLVQVSTAPTSLVLSWDPAYANGP----- 557
Qy 700 QSQYIQGYKILYR--SGANHG-ESDWLVEFVTRPAKNSVVPDLRKGVNYEIKARPFEN 756
| | : | | | | | : | | : | | |
Db 558 ---VQGYRLFCATFSGREQNIEVDGIVYR-----LEGLRKFTYSIRVLAYNR 603
Qy 757 EFQADSEIKFAKTLLEEAPSAPPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQYKVV 816
| | : | | | | | : | | : | | |
Db 604 YGPGVSSSEHTVTVLSDVPSAMPQVNSLEV--ANSRSIKVSWLPPPGTQNGFITGKIR 661
Qy 817 CLGNTRYHINKIVDGTFSVVPFLVPGIRYSVEVAASTGAGSGVKS-----EPQFIQL 871
| | : | | | | | : | | : | | |
Db 662 HRKTTTRGEL-ETLEPNLWYLTGLEKGSQSFQVAAVTNVTGTPSSDWYTAETPENDL 720
Qy 872 D-----AHGNPVPEDQVSLAQISDVVVKQPAFIAGIAGAAC----- 907
| | : | | | | | : | | : | | |
Db 721 DESQVPDQPSLHVRPLTTSIIMSWTPLNPNIVVRGYIIGYGVGSPYAEVTRVDSKQRY 780
Qy 908 -----WILMVFS-----IWLYRHRKKRN-----GLTSTYA 933
| | : | | | | | : | | : | | |
Db 781 YSIENLEPSSHYVLSKAFNNAEGVPLYESATTSQTPVDMSTPMLPPVGVQVAALVTHD 840
Qy 934 GIR-----KVPSFTFTPTVTVYQRGGEAVS-----SGGRPG-LLNI 967
| | : | | | | | : | | : | | |
Db 841 AVRYSWADNSVLKNQKTEVRFYTIWRVTSYSSASKYSADTTSLSHTVTGLKPNMYEF 900
Qy 968 SEPAAQWLAOTWPNNTGNHNDSCISCTAGNNGSDSNLTYSR---PADCIANYNNQLD 1024
| | : | | | | | : | | : | | |
Db 901 SVMYTKGRSSSTWSMTAH-----ATTYETAPTSAPKDLTVITREKPRAVIVSWQPPIE 954
Qy 1025 -----NKQTNLMLE---STVYGDVDSLKNINEMKT----FNSPNLKDGRFVN 1065
| | : | | | | | : | | : | | |
Db 955 ANGKIIDFILEYTLDKNLQDDWIMVTITGD-RLTHEILDNLDTAYFRIQARNAGLG 1013
Qy 1066 PSGQPTPYATTLQISNLSNNMNGSGDSEKHWKPLGQKQKEVAPVQYNIVEQNKLND 1125
| | : | | | | | : | | : | | |

Db 198 VTSTVTNARVLGSPPTPLVLRSDGVMGEYEPKIE-----LQFPETLPAAGSTVKL 247
Qy 282 KCEARGDPVPTVRWRKDDGELPKSRYEIRD-DHTLKIRKVTAGDMGSYTCVAENMVGAKE 340
Db 248 ECFALGNVPVQINWRRSDGMPFPPTKIKLRKFNGLBEPNQEDTGSYECIAENSRGKNV 307
Qy 341 ASATLTVQEPHFVVKPRDQVVALGRVTFQCEATGNPQAFWRREGSONLLFSYQPPQ 400
Db 308 ARGRLTYAKPYWVQLKDVETAVEDSLYWECRASGKPKPSYRWLKNGDALVL----- 360
Qy 401 SSSRFSVSQTGDLTITNVQRSDVGYIICQTLNVAGSIITKAYLEVTDIADRPVPIRQ 460
Db 361 -EERIQT-ENGALTITNLNVDGSMFQCIENKGLIYSSAELK---VLASAPD--FSRN 413
Qy 461 PVNQ--TVAVDGTFFVLSVATGSPVPTILWRKDGVLVSTQDSRIKLENGVLQIRYAKLG 518
Db 414 PMKMKIQQVQGLVILDCPKSPRALSFWKGDVVRQ-ARISLNDGGLKIMNVTKA 472
Qy 519 DTGRYTCIASTPSGEATWSAYIEVQFVGVPVQPPRPTDNLIPS----- 562
Db 473 DAGIYTCAENQFGKANGTQLVVT-----PTRIILAPSNMDVAVGESIILPCQV 523
Qy 563 ----- 562
Db 524 QHDPLLDIMFAWYFNGTLTDFKKGDSHFKEVGGSSGDLMIQNLKHSKGYKVMQTV 583
Qy 563 -----APSKPE---VTDVSRNTVTLVSWQPNLNSGATPTSIIIEAFSHASGSS 606
Db 584 DSVSSAAELIVRGSPGPPENVKVDEITDTAQLSWTEGDSHSPVISYAVQARTPFS-VG 642
Qy 607 WQ---TVAENV--KTETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTDVLPTS 661
Db 643 WQNVRTVPEAIDGKTRTATVVELNPWVEYEFVVASNKIGGGEPSLPSEKVRTEAAPE- 701
Qy 662 QGVDPKQVQRELGNVHLHNPVTLSSSSIEVHWTVDDQSQYIQ-----GYKILYRPSGA 716
Db 702 --VAPSEVSGGG-----SRSELVITW--DPVPEELQNGGGFGYVAVRPLGV 745
Qy 717 NHGESDWLVFEVTRP-----AKNSVVPDLRKGVNYEIKARPFNEFGADSEIKFAKT 770
Db 746 ----TWIQTVTSPDNPRIVFRNESIVP----FSPTEYKGVYNNKGEPPSPVTTVFS 797
Qy 771 LEEAPSAPPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQYEVK--WCLGNTRYHINK 828
Db 798 AEEETPVAPS--HISAHLSSSEIEVSWNTIPWKLNSGRLGLGYEVRYWNGGEESSSKV 855
Qy 829 TVDGSFVSVPFLVPGIRYSVEVAASTGAGSG 861
Db 856 KVAGNQTSAVLRLGLKSNLAYTAVRAYNTAGAG 888

RESULT 14

A53449

plasmacytoma-associated neuronal glycoprotein PANG - mouse

C;Species: Mus musculus (house mouse)

C;Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 24-Sep-1999

C;Accession: A53449

R;Connelly, M.A.; Grady, R.C.; Mushinski, J.F.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 91, 1337-1341, 1994

A;Title: PANG, a gene encoding a neuronal glycoprotein, is ectopically activated by intr

A;Reference number: A53449; MUID:94151325

A;Accession: A53449

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1028 <CON>

A;Cross-references: GB:L01991; NID:g200056; PIDN:AAA17403.1; PID:g200057

C;Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

C;Keywords: glycoprotein

Query Match 8.0%; Score 695.5; DB 2; Length 1028;

Best Local Similarity 25.5%; Pred. No. 2.4e-25;

Matches 238; Conservative 125; Mismatches 361; Indels 209; Gaps 28;

Qy 68 PRIVEHPSDLIV---SKGEPATLNCKAEGRPPTTIEWYKGERVETDKDDPRSHRMLPS 124
Db 26 PVFIKEPSNSIFPVDSDEKKITLNCCEARGNPSPHYRWQLNGSDIDTSLD----HRYKLNG 81
Qy 125 GSLFFLRIVHGRKSRPDEGVYVCAVNYLGEAVSHNASLEVAIRDDFRQNPSPDMVAVG 184
Db 82 GNL----IVINPNRNDTGSYQCFATNSLGTIVSREAKLQFAYLENFKTRMRSTVSVRG 137
Qy 185 EPAVMCOPPRGHPEPTISWKKDGSL---DDKDERITIRGGKLMITYTRKSDAGKYVCV 241
Db 138 QGVVLCGPPPHSGELSYAWFNEYPSPVEEDSRRLVSQETHLYIAKVEPSDVGNTCV 197
Qy 242 GTNMV-----GERESEVAELTVLERPSFVKRPSNLAVTVDDSAEF 281
Db 198 VTSTVTNARVLGSPPTPLVLRSDGVMGEYEPKIE-----VQFPETLPAAGSTVRL 247
Qy 282 KCEARGDPVPTVRWRKDDGELPKSRYEIRD-DHTLKIRKVTAGDMGSYTCVAENMVGAKE 340
Db 248 ECFALGNVPVQINWRRSDGMPFPNKIKLRKFNGLBEPNQEDTGSYECIAENSRGKNV 307
Qy 341 ASATLTVQEPHFVVKPRDQVVALGRVTFQCEATGNPQAFWRREGSONLLFSYQPPQ 400
Db 308 ARGRLTYAKPYWVQLKDVETAVEDSLYWECRASGKPKPSYRWLKNGDALVL----- 360
Qy 401 SSSRFSVSQTGDLTITNVQRSDVGYIICQTLNVAGSIITKAYLEVTDIADRPVPIRQ 460
Db 361 -EERIQT-ENGALTITNLNVDGSMFQCIENKGLIYSSAELK---VVASAPD--FSRN 413
Qy 461 PVNQTVAVD--GTFFVLSVATGSPVPTILWRKDGVLVSTQDSRIKLENGVLQIRYAKLG 518
Db 414 PMKMKIQQVQGLVILDCPKSPRALSFWKGDVVRQ-ARVSLNDGGLKIMNVTKA 472
Qy 519 DTGRYTCIASTPSGEATWSAYIEVQFVGVPVQPPRPTDNLIPS----- 562
Db 473 DAGIYTCAENQFGKANGTQLVVT-----PTRIILAPSNMDVAVGESIILPCQV 523
Qy 563 ----- 562
Db 524 QHDPLLDIMFAWYFNGALTDFKKGDSHFKEVGGSSGDLMIQNLKHSKGYKVMQTV 583
Qy 563 -----APSKPE---VTDVSRNTVTLVSWQPNLNSGATPTSIIIEAFSHASGSS 606
Db 584 DSVSSAAELIVRGSPGPPENVKVDEITDTAQLSWTEGDSHSPVISYAVQARTPFS-VG 642
Qy 607 WQ---TVAENV--KTETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTDVLPTS 661
Db 643 WQSVRTVPEVIDGKTHATVVELNPWVEYEFVVASNKIGGGEPSLPSEKVRTEAAPE- 701
Qy 662 QGVDPKQVQRELGNVHLHNPVTLSSSSIEVHWTVDDQSQYIQ-----GYKILYRPSGA 716
Db 702 --IAPSEVSGGG-----SRSELVITW--DPVPEELQNGGGFGYVAVRPLGV 745
Qy 717 NHGESDWLVFEVTRP-----AKNSVVPDLRKGVNYEIKARPFNEFGADSEIKFAKT 770
Db 746 ----TWIQTVTSPDNPRIVFRNESIVP----FSPTEYKGVYNNKGEPPSPVTTVFS 797
Qy 771 LEEAPSAPPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQYEVK--WCLGNTRYHINK 828
Db 798 AEEETPVAPS--HISAHLSSSEIEVSWNTIPWKLNSGRLGLGYEVRYWNGGEESSRKV 855
Qy 829 TVDGSFVSVPFLVPGIRYSVEVAASTGAGSG 861
Db 856 KVAGNQTSAVLRLGLKSNLAYTAVRAYNSAGAG 888

RESULT 15

A39640

neural cell adhesion molecule Nr-CAM precursor - chicken

C;Species: Gallus gallus (chicken)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: A39640; S16451

Qy 729 RTPAKNSVVIPDLRK-----GVNIEIKARPFNEFQADSEIKFAKYLEAPSAPPQ 780
 ||| : : ||| : : : : : |
 Db 772 ----TSVVIVANSKYIVSGTPTVPEIKVQALNDIGYAPRPSVEIGHSGFIDPWVRC 826

Search completed: January 22, 2001, 12:27:16
Job time: 2113 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:28:41 ; Search time 162.41 Seconds
(without alignments)
328.290 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKWKHPVPLVMISLSPN.....VLGGYERGEDNNEELEETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	764.5	8.8	1260	1	CAML_MOUSE	P11627 mus musculus
2	761	8.7	1257	1	CAML_HUMAN	P32004 homo sapien
3	760	8.7	1493	1	NEO1_MOUSE	P97798 mus musculus
4	757.5	8.7	1443	1	NEO1_CHICK	Q90610 gallus gall
5	755	8.7	2012	1	DSCA_HUMAN	O60469 homo sapien
6	746.5	8.6	1377	1	NEO1_RAT	P97603 rattus norv
7	739.5	8.5	1259	1	CAML_RAT	Q05695 rattus norv
8	725.5	8.3	1461	1	NEO1_HUMAN	Q92859 homo sapien
9	701.5	8.0	1284	1	NRCA_CHICK	P35331 gallus gall
10	679.5	7.8	1447	1	DCC_MOUSE	P70211 mus musculus
11	667.5	7.7	1447	1	DCC_HUMAN	P43146 homo sapien
12	653.5	7.5	1040	1	AXO1_HUMAN	Q02246 homo sapien
13	652	7.5	1040	1	AXO1_RAT	P22063 rattus norv
14	633.5	7.3	1036	1	AXO1_CHICK	P28685 gallus gall
15	627	7.2	1239	1	NRG_DROME	P20241 drosophila
16	596.5	6.8	1018	1	CONT_HUMAN	Q12860 homo sapien
17	589.5	6.8	1010	1	CONT_CHICK	P14781 gallus gall
18	584	6.7	1912	1	PTPD_HUMAN	P23468 homo sapien
19	583	6.7	2029	1	LAR_DROME	P16621 drosophila
20	573	6.6	1020	1	CONT_MOUSE	P12960 mus musculus
21	549	6.3	1266	1	NGCA_CHICK	Q03696 gallus gall
22	549	6.3	1897	1	PTPF_HUMAN	P10586 homo sapien
23	505	5.8	1070	1	PTK7_HUMAN	Q13308 homo sapien
24	500.5	5.7	3707	1	PGBM_MOUSE	Q05793 mus musculus
25	500	5.7	1091	1	NCA1_CHICK	P13590 gallus gall
26	497.5	5.7	1115	1	NCA1_MOUSE	P13595 mus musculus
27	495	5.7	4393	1	PGBM_HUMAN	P98160 homo sapien
28	490	5.6	1051	1	PTK7_CHICK	Q91048 gallus gall
29	483.5	5.5	853	1	NCA1_BOVIN	P31836 bos taurus
30	478.5	5.5	848	1	NCA1_HUMAN	P13591 homo sapien
31	466	5.3	761	1	NCA2_HUMAN	P13592 homo sapien
32	465	5.3	858	1	NCA1_RAT	P13596 rattus norv
33	460	5.3	1092	1	NCA2_XENLA	P36335 xenopus lae

34	450	5.2	837	1	NCM2_MOUSE	O35136 mus musculus
35	449.5	5.2	1088	1	NCA1_XENLA	P16170 xenopus lae
36	448.5	5.1	725	1	NCA2_MOUSE	P13594 mus musculus
37	426.5	4.9	837	1	NCM2_HUMAN	Q15394 homo sapien
38	422.5	4.8	811	1	FS22_DROME	P34083 drosophila
39	422.5	4.8	873	1	FS21_DROME	P34082 drosophila
40	415.5	4.8	1906	1	KMLS_CHICK	P11799 gallus gall
41	401.5	4.6	2481	1	UN52_CAEEL	Q06561 caenorhabdi
42	398.5	4.6	1913	1	KMLS_HUMAN	Q15746 homo sapien
43	390.5	4.5	1274	1	MYPC_HUMAN	Q14896 homo sapien
44	387.5	4.4	1270	1	MYPC_MOUSE	O70468 mus musculus
45	379	4.3	1271	1	MYPC_CHICK	Q90688 gallus gall

ALIGNMENTS

RESULT 1

CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-BRAIN;
RX MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplov D., Frueh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
RT superfamily with binding domains similar to fibronectin."
RL Nature 334:701-703(1988).
CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12875; CAA31368.1; -.
DR PIR; S05479; S05479.
DR HSSP; P20241; 1CFB.
DR MGD; MGI:96721; L1CAM.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
DR PRINTS; PR00014; FNTYPEIII.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Signal; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1124 1146 POTENTIAL.
FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 50 120 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 150 215 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 256 318 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 346 410 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 440 503 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 531 599 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 827 896 FIBRONECTIN TYPE-III.
FT DOMAIN 932 994 FIBRONECTIN TYPE-III.
FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III.
FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1180 1183 MISSING (IN SHORT ISOFORM)
FT (BY SIMILARITY).
SQ SEQUENCE 1260 AA; 140968 MW; 22BE5701CE2A538 CRC64;

Query Match 8.8%; Score 764.5; DB 1; Length 1260;
Best Local Similarity 23.3%; Pred. No. 5.4e-27;
Matches 307; Conservative 177; Mismatches 510; Indels 325; Gaps 51;

Qy 56 YTGSRRLQEDFPPIRIVEH-PSDLIVSKGEPATLNCAGRPPTTIEWYKGGSERVETDK-- 112
| | : : | | | | : | : : | | | | : | : :
Db 26 YKGHHVLE---PPVITEQSPRLVVPPTDDISLCKEARGRPQVEFRWTKDGHFKPKKEEL 82
Qy 113 ----DDPRSHRLPSGSLFFLRIVHGRKSRDEGVYVCARNYLGEAVSHNASLEVAI 167
| | : : | | | | : | : : | | | | : | : :
Db 83 GVVVHEAPYSGSFTIEGNNNSFAQRF-----QGIYRCYASNKLTAMSH---EIQL 129
Qy 168 LRDDFRQNPDS---VMVAVGEPVMECPQGHPEPTISWKKDGLD--KDERITI- 220
| | : : | | | | : | : : | | | | : | : :
Db 130 VABGAPKPKETVKPVEVEEGSVVLPNCNPPSAAPPRIYWM--NSKIFDIQDERVSMG 187
221 RGGKLMITYTRKSD-AGKYVCVTNMVGER--ESEVAELLVLERPSFVKRPSNLAVTVD 276
| | : : | | | | : | : : | | | | : | : :
Db 188 QNGDLYFANVLTSNHSYDIC-NAHPFGTRTIQKEPIDLVRKPTNSMIDRKLPLEFTN 246
Qy 277 DSAB-----FKCEARGDPVPTVWRKDDGELPKSRYEIRDDH--TLKIRKVTAGD 324
| : | | | | | : | : | : | : | : | :
Db 247 SSSRLVALQGQSLILEICAEFPPTPIKWLHPSDPMPTDRV-IYQNHKTLQLLNVEED 305
Qy 325 MGSYTCVAENMVGAESATLTQVEPPHFVVKPRDQVVALGRTVTQCEATGNPQPAIFW 384
| | | : | : | : | : | : | : | : | : | :
Db 306 DGEYTCIAENSLGSARHAYYTVTAAPYALQKPKQSHLYGPGETARLDQVQGRQPPEITW 365
Qy 385 RREGSQNLFSYQPPQSSSRFSVSQGTDLITNVQSRDVGYYICQTLNVAGSIITKAYLE 444
| | : : | : | : | : | : | : | : | :
Db 366 RING-----MSMETVKNQDKYRIEQ-GSLILSNVQPTDMVTQCEARNQHGILLANAYIY 419
Qy 445 VTDVIADRPPIVIRQGPVNT-VAVDG-TFVLSCAVIGSPVPTILWRKDGVLSTQDSRI 502
| : | : | : | : | : | : | : | : | :
Db 420 VVQL----PARILTKD--NQTYMAVEGSTAYLLKAPGAPVPSVQWLDEEGTIVLQDERF 473
Qy 503 KQLENGVLQIRYAKLGTGRYTCIASTPSGEATWSAYIEVQFGVPVQPPRPT----- 555
| | | : | | | : | : | : | : | : | : | : |

Db 474 PPYANGTSLIRDLQANDTGRYFQCAANDQNNVITLANLQVKEATQITQGRPSAIEKKGAR 533
Qy 556 -----DPNL----- 559
| | : |
Db 534 VITFCQASFDPSLQASITWRGDGRDLQERGSDSKYFIEDGKLVQSLDYSQGNYSVAS 593
Qy 560 -----IPSAPSKPEVTD---VSRNTVTLSWQPNLNSGATPTSYIIIE-AF 599
| : | : | : | : | : | : | : | : | :
Db 594 TELDEVESRAQLLVGSPGPVPHLELSRHLKQSQVHLSWSPAEDHNSPIEKYDIEFED 653
Qy 600 SHASGSSWQTVAEVNVKTETSAIKGLKPNAYILFLVRAANAYISDPSQISDPVKTDVLP 659
| : | : | : | : | : | : | : | : | :
Db 654 KEMAPEKWFSLGKVPNGQSTTLKLSPVHYTFRTVTAINKYGPGEPSVSESVVTPPEAP 713
Qy 660 TSQGVHDHKVQRELGNVHLHNPVTLSSSSIEVHWTVDQSQYIQYKILYRPSGANHG 719
| | : | : | : | : | : | : | : | : | :
Db 714 EKNPVDVRGEGNETNNMVI-----TWKPLRW-MDWNAPQIQ-YRQWRPQKG--- 758
Qy 720 BSDLVLFVVRTPAKNSVVPDLRKGVNYEIKARFFNEFQADSEIKFAKTEEAPSAPP 779
| : | : | : | : | : | : | : | : | :
Db 759 QETWRKQTVSDP---FLVVSNTSTFVPYIEKVQAVNNQKQGPQVITIGSGEDYQVQSP 815
Qy 780 --QGVTVSKNDGNGTAILVSWQPPEDTQNGMVQEYKV--WCLGNTRY---HINKT--- 829
| : | : | : | : | : | : | : | : | :
Db 816 ELEDITIF---NSSTVLVRWPVDLAQVKGHLKGYNVITYWKGSRKHSKRHIKSHIV 871
Qy 830 VDGSTFSSVPIFLVPGIRYSVEVAASTGAGSGVKSE-----PQFIQLDAHG-- 875
| : | : | : | : | : | : | : | : | :
Db 872 VPANTTSAILSGLRPSYSHVEVQAFNGRGLGPASEWFTSTPEGVPHGHEALHLEQSDT 931
Qy 876 -----NPVSPEDQVSLAQQISDVVKQPAFIAGIAGAACWIILM 912
| : | : | : | : | : | : | : | : | :
Db 932 SLLLHWQPLSHNGVLTGYLLSYHVEGESKEQLFFNLSD----- 971
Qy 913 VFSIWLYRHRKRNGLTSTYAGIRKVSFTFTPTVTYQRG--GEAV-----SSGGRPG 963
| : | : | : | : | : | : | : | : | :
Db 972 -----PELRTHNLNLNPDQ---YRFQLQATQGGGPEAIVREGGTALFGKPD 1019
Qy 964 LLNISEPAAQWPLADTW-PNTGNHNDSCISCTAGNGNSDNLTYSRPADCIANYNN- 1021
| | : : | : | : | : | : | : | : | :
Db 1020 FGNISATAGENYSVSVVPRKG---QCNFRFHILFKALPEGKVSQDPQVQYVSYNQS 1075
Qy 1022 -----QLDNKQTNLMPESTVYGDVDSL------KINEMKTFNS----- 1055
| | : : | : | : | : | : | : | : | :
Db 1076 SYTQWNLPDQTKYIEHLIKEVLLHLDVTNGTGPVRVSTTGSFASGEWFIAFVSAAIL 1135
Qy 1056 -----PNLKDGRF-----VNPSCQPTPIYATQLIQSNLNNMNGSGDGS 1095
| : | : | : | : | : | : | : | : | :
Db 1136 LLLILLILCFIKRSKGGKYSVKDKEDTQVDESEARPMKDETFGYRSLESDNEEKAFSGSQ 1195
Qy 1096 EK---HWKPLGQKQEV-----APVQYN---IVEQNKLNKDYRA---NDTVPTPIYN 1139
| | | | | : | : | : | : | : | : | :
Db 1196 PSLNGDIKPLGSDSLADYGGSDVQFNEGDSFGIQYSGKKEEAAGNDSGATSPIN 1254

RESULT 2

CAML_HUMAN

ID CAML_HUMAN. STANDARD; PRT; 1257 AA.
AC P32004;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN LICAM OR CAML1 OR MIC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92031698; PubMed=1932117;
RA Kobayashi M., Miura M., Asou H., Uyemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of

- RT different origin.";
RL Biochim. Biophys. Acta 1090:238-240(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Rosenthal A., Coutelle O., Drescher B.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92329299; PubMed=1627459;
RA Reid R.A., Hemperly J.J.;
RT "Variants of human L1 cell adhesion molecule arise through alternate
RT splicing of RNA.";
RL J. Mol. Neurosci. 3:127-135(1992).
RN [4]
RP SEQUENCE OF 353-1176 FROM N.A.
RX MEDLINE=92020233; PubMed=1923824;
RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
RT "PCR walking from microdissection clone M54 identifies three exons
RT from the human gene for the neural cell adhesion molecule L1
RT (CAM-L1).";
RL Nucleic Acids Res. 19:5395-5401(1991).
RN [5]
RP SEQUENCE OF 332-371 FROM N.A.
RX MEDLINE=90353957; PubMed=2387585;
RA Djabali M., Mattei M.-G., Nguyen C., Roux D., Demengeot J.,
RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
RT "The gene encoding L1, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [6]
RP SEQUENCE OF 1030-1257 FROM N.A.
RX MEDLINE=91132183; PubMed=1993895;
RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
RA Stallcup W.B.;
RT "Isolation and sequence of partial cDNA clones of human L1: homology
RT of human and rodent L1 in the cytoplasmic region.";
RL J. Neurochem. 56:797-804(1991).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE=88298876; PubMed=3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule L1.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP VARIANT HSAS TYR-264.
RX MEDLINE=94004956; PubMed=8401576;
RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnai D.;
RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANT HSAS/MASA LEU-1194.
RX MEDLINE=95187172; PubMed=7881431;
RA Fransen E., Schrandt-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the L1CAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE=95004608; PubMed=7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metzenberg A., Ionasescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the L1 gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE=95004609; PubMed=7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boulle K.,
RA Reyniers E., Korn B., Poustka A., Wilson G., Schrandt-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT L1CAM.";
RL Nat. Genet. 7:408-413(1994).
RN [12]
RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
RX MEDLINE=95282776; PubMed=7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Wadelius C., Kenwick S.;
RT "New domains of neural cell-adhesion molecule L1 implicated in
RT X-linked hydrocephalus and MASA syndrome.";
RL Am. J. Hum. Genet. 56:1304-1314(1995).
RN [13]
RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
RX MEDLINE=96153146; PubMed=8556302;
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, L1.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]
RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Frys J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in L1-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; V-439--T-443 DEL; C-784 AND L-936--L-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).
RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE=98112489; PubMed=9452110;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
RL Hum. Mutat. Suppl. 1:S284-S287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE=98415726; PubMed=9744477;
RA Saugier-Verber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
RT "Identification of novel L1CAM mutations using fluorescence-assisted
RT mismatch analysis.";

```
Qy      14 LLSLSPNHLFLAQLIPDPEDVERGNDHGPTIPTSDNDNSLGYYTGSRLROEDFPFRIVEH 73
       || ||           || ||: | :   | : | |           | | | |
Db      11 LLCLSP-----CLLIQIPEEYEGHHVMEPPVIT-----EQSPRRLVVF 48
               | | | | | | | | | | | | | | | | | | | | | |
Qy      74 PSDLIVSKGEPAATNLCKAEGRPTPIIEWYKGGERVETDKD-----DPRSHRMLLPSGS 126
       |:| |         |:| | | | | | | | | | | | | | | | | |
Db      49 PTDDI-----SLKCEASGKEPVQFWRTRGVGHFKPKELGVTVYQSPHSGSFITIGN 101
               | | | | | | | | | | | | | | | | | | | | | |
Qy      127 LFFLRIVHGKRSPDEGYVVCVARNYLGEAVSHNASLEVALRDDFRONPSD---VMVA 182
        |          :: |:| | | | | | | | | | | | | | | | | |
Db      102 SNF-----AQRFQGIYRCFASNKLGTAMSH-----EIRLMAGAPKWPKETVPKVEVE 149
               | | | | | | | | | | | | | | | | | | | | | |
Qy      183 VGEPAMVECQPGRGHPEPTISWKKDGSPDLLDKDERITI--RGGKLMITYTRKSD-AGKYVC 240
        || | : | | | | | | | | | | | | | | | | | | | | | |
Db      150 EGESVVLPCNPPPSAEPLRIYWMSNKILHIKIQDERVTMGQNGNLYFANVLTSDNHSDYIC 209
               | | | | | | | | | | | | | | | | | | | | | |
Qy      241 ----VGTMNGVERESEVAELTVLERPSFVKRPSNLAVTVDDSAE-----FKCEAR 286
        || : :: | : | | | | | | | | | | | | | | | | | |
Db      210 HAHFGPGRITIIQKEP--IDLRVKATNSMDRKPRLLPFTNSSSHLVALQGQPLVLCIEAE 267
               | | | | | | | | | | | | | | | | | | | | | |
Qy      287 GDPVPTVRWRKDDGELPKSRYEIRD-DHTLKIRKVTAGDMGSYTCVAENMVGKAESATL 345
        | | | | | | | | | | | | | | | | | | | | | | | |
Db      268 GFPTPTIKWLRSGLPMADRVTYIQHNKNTQLLKVGEEDGEYRCLAENSLGARHAHYTV 327
               | | | | | | | | | | | | | | | | | | | | | |
Qy      346 TVQEPPHFVVKPRDQVVALGRVTVFQCEATGNQPAIFWRREGSQNLLFSYQPPQSSSRF 405
        || : :: | | | | | | | | | | | | | | | | | | | | | |
Db      328 TVEAAPYLWHKPOSHLYGPGETARLDCCVQGRQPEVTVWRING----IPVEELAKDQKY 382
               | | | | | | | | | | | | | | | | | | | | | |
Qy      406 SVSQTGDLTITNWQRSDVGYYICTLNVAGSIITKAYLEVTVDIADRPPPVIRQGPVNQT 465
        || | | | | | | | | | | | | | | | | | | | | | |
```

[illegible]

```

RESULT      3
NEO1_MOUSE
ID  NEO1_MOUSE      STANDARD;      PRT;   1493 AA.
AC  P97798;
DT  01-OCT-2000 (Rel. 40, Created)
DT  01-OCT-2000 (Rel. 40, Last sequence update)
DT  01-OCT-2000 (Rel. 40, Last annotation update)
DE  NEOGENIN PRECURSOR.
GN  NEO1 OR NGN.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN  [1]
RC  SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC  TISSUE-BRAIN;
RX  MEDLINE=97407661; PubMed=9264410;
RA  Keeling S.L., Gad J.M., Cooper H.M.;
RT  "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RT  expressed widely in the adult mouse and during embryogenesis.";
RL  Oncogene 15:691-700(1997).
CC  -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC      TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC      DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC      MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC  -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC  -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4

```

```

Db      123 SNVHSHKHNKPDEGFYQCVATVDNLGTVSRKAKTLVAGL-PRFTSQPEPSSVYVGNLSAI 181
Qy      189 MECQPQRPHPEPTISWKKDGSPLDDKERITIRGGKLIMITYRKSDAGKYVCVTGNMVG 248
        :|: |::||| || | : : |:|: : ||| |: 
Db      182 LNCE-VNADLVFPVRWNQNRQLLLDDRIKLPSTGLVISNATGEKGGLRRCIVESGGPP 240
Qy      249 RESEVAELTVLERPS-----FVKRPSNLAVTVDDSAEFKCARGDPVPTVRWRKDDGEL 302
        :|: ||| |:| :|: ||::| | | | | |:| 
Db      241 KFSDEAELKVLDQPEEIVDLVFLMRPSSMMKVITGQSAYLPCVVSGLPAPVVRWMKNEEVL 300
Qy      303 ---PKSRYEIRDHDHTLKIRKVITAGDMGSYTCVAENMVGKAESATLTVQEPPHFVVKPRD 359
        :|: |::||| |:| |:|:| | | ||| |:| : 
Db      301 DTESSGRLVLLAGGCLEISDVTEDDAGTYFCTIADNGMKTVEAQALTVQVPGLKQPAN 360
Qy      360 QVALGRVTTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGDLTITNVQ 419
        :|: ||| |:| :|: |:| :| | | | |:| : 
Db      361 IYAHEMDIVFECEVTGKPTPTPVKKWKNGDVVI-----PSDNFKVIKEHNLQVLGLV 412
Qy      420 RSDVGYICQLNLVAGSIITKAYLEV---DVIADRPPIRVGPVNQIVAVDGTFFVLSCV 477
        :|| |:| |:| |:| :|: | | | | | | 
Db      413 KSDEGYPQCIANDVGNQAQAQAILILEHDAIPTLPP-----TSLTATTDLHLAP 463
Qy      478 ATGPSVPTTILWRKDGVLSTQDSRIKQLENGVLQIRYAKLGDTGRYTIASTPSGE-ATW 536
        || |:|:| | | | | | | | | | | | | | 
Db      464 ATTGPLSPAPRDVVASLVST-----RFIKL----TWRTPASDPHGDNLT 504
Qy      537 SAYIEVQEFQVPVQPPRPDTPNLIPSPAPSKPEVTDVSRNTVTLSWQPNLNSGATPTSII 596
        |:| :|: |:| :| | | | | | | | | | 
Db      505 SVFYTKB-GVDRERVENT-----SQPGEMQVT----- 530
Qy      597 EAFSHASGSSWQTVAENVKTETSAIKGLKNPAIYLFVRAANAYGISDPSQISDPVKTOD 656
        |:| |:| :|:| |:| |:| |:| |:| |:| 
Db      531 -----IQNLMPATVYIFVKMAQNKHG-SGESSAPLRVETQP 565
Qy      657 VLPTSQGVDHQVQRELGNALVHLNPTVLSSSSIEVHW-TVDQQSQYIQGYKILYRPSG 715
        :|: |:| | :|:| | | | | | | | | | 
Db      566 EV-----QLPGPAPNIRAYATSPT----SITVTWETPLSGNGEIQNKLYIMEKG 611
Qy      716 ANHGESDWLFEVTRTAPKNSVVIDLRKGVNYEIKARPFNEFGADSEIKFAKTEEAP 775
        :| | | :|:| |:| |:| :|:| :|:| |:| 
Db      612 TDK-EQDIDV-----SSHYSITINGLKXYTEYSFRVVAYNKHGPGVSTQDVAVRTLSDVP 664
Qy      776 SAPPGQVTVSKNDGNGTAILVSWQPPPEDTQGMVQYEYKWCLGNTRYHINKT-VDGST 834
        || || ::| | :|:| ||| ||| :||:| :|:| |:| 
Db      665 SAAPONLSLEVR--NSKSIVIHWPQSPSTTQNGQITGYKIRYKASKRSQVTELTVTGTQ 722
Qy      835 FSVVIPFLVPJIRYSVEVAASTGAGSGVKSE-----PQFIQLDAHGNPV 878
        |:| |:| |:| |:| |:| |:| :|:| |:| |:| 
Db      723 LSQLEGLDRGETEYNFRVAALTVMGTGPATDWLSAETFESDLDETROPV-VPSSLHVRL 781
Qy      879 -----SPEDQVSLAQQISDVVKQPAFIAGIGAACWIILMVFSIWLYRHRRKKNGLT 929
        ||:| :|:| |:| |:| |:| |:| |:| 
Db      782 VTSIVVSWTPPENQ-----NIIVRGYAIYGIGSPAQTIKVD---YKQR----- 823
Qy      930 STYAGIRKV-PSFTFTPTV-TYQRGGEAV-----SSGGRP-----GLLNISEP-AAQPW 975
        |:| |:| |:| |:| |:| |:| |:| |:| |:| 
Db      824 --YTTIENLDPSHHVITLKFANNVGEIGPLYESAIVRPHDTSSEVDFVINAPYTPVPD 881
Qy      976 LADTWPTNG-----NNHNDCSISCTAGNNSDSNLTTISRADCIANYNQLDNKQTNL 1030
        |:| |:| :|:| |:| |:| |:| |:| |:| |:| 
Db      882 PTPMPPVGVQASILSHDTIRITW-----ADNSLPKPKQKITD--SRYITV-RWKTN- 929
Qy      1031 MLPESTVYGDVDSLKNKINEMKTFNSPN-----LKDGRFVNPSGQPTPYATTQLIQSN 1082
        :|: |:| :|:| :|:| |:| |:| |:| |:| |:| 
Db      930 -IPANTKYKNAN-ATTLSYLVTGLKNPWLIEFSVMVTKGRSSWTSMTAGATFELVPTS 987
Qy      1083 LSNMNMGSGSGSEX---HWKPLGQQKQEV-----APVQYNIVEQ---NKLN 1123
        |:| |:| :|:| |:| |:| |:| |:| |:| |:| 
Db      988 PKPDVTVVSKEGKPTIIVNQPPSEANGKITGYIIYYSTDVNAIEHDWVIEPVVGNRLT 1047
Qy      1124 KDYRANDTVPTTIPYNQSYDQNTGG-----SYNSSDR--GSSTSGSGQHKKG 1168

```



```
Qy 1479 DLPPPPV-----PPPAIKSPTAQSKTQLEVR---PVVVVKLPSPMDARTDRSSDRKGS 1527
      || :|      ::||      | ||||| :| :| :| :|
Db 1360 TLPSTPLLTQQAPSHPHSVK--TASIGTLGRTRPPMPVVVPSAPDVQ-ETTRMLEDSSES 1416
      || :|
Qy 1528 SYKGREV 1534
      || :|
Db 1417 SYEPDEL 1423
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/annou>

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Db 1316 AVKWMKDSNGTSLVITDGRSIFSNFSFIIRTVKAEDSGYISCIANNWGSDEILNLQ 1375
Qy 1025 -----NKQTNLMPESTVYGD-----VDLSKNINEM-----KT 1052
Db 1376 VQVPPDQRLTVSKTITSSITLSWLPDNGSSIRGIYLIQYSEDNSEQWGSFPIPSERS 1435
Qy 1053 FNSPNLKDGRF-----YNPS-----GQPTPYATTQLIQSNLSNMNM-- 1089
Db 1436 YRLNKKCGTWYKFTLTQNG/GPGRISEIEAKTLGKEPQFSKEQELFASINTTRVRLN 1495
Qy 1090 --GSGDSG-----EKHWKPLEQQQOEAVPVQYINEQVKNLNDY----- 1126
Db 1496 LIGWNGGCPITSFTLEYRPF-----TTVWTAQRTSLSKSYLYLDQEATWYELQM 1548
Qy 1127 -----RAN-----DIVPPTIPYQSDQNTGGSYNSSDRGSSSTGSQGHK- 1166
Db 1549 RVCNSAGCAEQANFATLNDSTIIPPLI---KSVVQN-----EEGLTT---NEGLAM 1595
Qy 1167 -----KGARTPKVPR-QGGMWADLLPPPAHPPPHSSEYN 1203
Db 1596 LVTISCLVGVLLLVLLVRRRREQLRKLDAKSLAELM-----MSKNRTSD 1647
Qy 1204 ISVDESIDQEMPCVPVPMRLQQDELEEEEDERGPP---TPVGAASSPAVSYSH--- 1257
Db 1648 TLSKQQTLRMHIDIPRAQLLIEERDMETIDDRSTVLLTDADFGAAKQSLVTHTVH 1707
Qy 1258 -QSTALTPSPQELQPMQLQDCPEETGHMQHPDRRRQVPSPPPP-----PRP 1304
Db 1708 YQSVSQAT-----GPLVDVSDARPQ---TNPTTRNAKAGPIARNRYASQWTLNRPHP 1757
Qy 1305 ISPHTYGYISGPLVSDMDTAP-----EEEEDEADMEVAKMOTRRLLRGLEQTASS 1358
Db 1758 TISAT-----LTTDWRLPTPRAAGSVKEDSYSVSPSQDTDR-----ARSS 1800
Qy 1359 VGDLE--SSVTGSMINGWSASEEDNISGRSSVSSDGSFPTDADFAQAAAAEYAGL 1416
Db 1801 MVSTESASSTYELARAYEHAKMEEQLRHAFTITE---CPISDTSSQLTAGTNEYTDS 1857
Qy 1417 KVARROMQDAAGRRHFHASQCPRPTSPVSTDSNMSAAVQKTRPAKCLKHQPGLRRRETY 1476
Db 1858 LTSSTPSE--SGICRTAS---PPKPDGGRVMMAVPAHRPG-DLIHLPPYLRMDPL 1910
Qy 1477 TDDLPPPPVPPPAIKSPTAQ-SKTQLEVRPVVVKPLSMDARTDRSSDRKGSSYK----- 1530
Db 1911 LNRGGPGTSRDLGLQACLEPQKSRTRLKRPVLEPIP-MEAASSASTREGSQWPGAVA 1969
Qy 1531 ---GREVLDRQVDMRTNPGDPREAEQONDGKG--RGNAAKRDLPPAKTHLI 1580
Db 1970 TLPQREGAELGQAAMSS-----SQESLLDSRGLKGNV-----PYAKSYTL 2011

RESULT 6

NEO1_RAT

ID NEO1_RAT STANDARD; PRT: 1377 AA.

AC P97603;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NEOGENIN PRECURSOR (FRAGMENT).

GN NEO1 OR NGN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=97015074; PubMed=8861902;

RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-Y.,

RA Culotti J.G., Tessier-Lavigne M.;

RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";

RL Cell 87:175-185(1996).

CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE

CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U68726; AAB41100.1; -.

DR HSP; P56276; 1TLK.

DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00041; fn3; 6.

DR PFAM; PF00047; ig; 4.

KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal.

FT NON_TER 1 1

FT SIGNAL <1 2 POTENTIAL.

FT CHAIN 3 1377 NEOGENIN.

FT DOMAIN 3 1074 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1075 1095 POTENTIAL.

FT DOMAIN 1096 1377 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 36 105 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 135 197 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 232 296 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 324 386 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 405 502 FIBRONECTIN TYPE-III.

FT DOMAIN 505 598 FIBRONECTIN TYPE-III.

FT DOMAIN 599 698 FIBRONECTIN TYPE-III.

FT DOMAIN 704 798 FIBRONECTIN TYPE-III.

FT DOMAIN 819 919 FIBRONECTIN TYPE-III.

FT DOMAIN 920 1021 FIBRONECTIN TYPE-III.

FT DOMAIN 1087 1090 POLY-VAL.

FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 878 878 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1377 AA; 150637 MW; E514ED8ABD1A63A9 CRC64;

Query Match 8.6%; Score 746.5; DB 1; Length 1377;

Best Local Similarity 22.6%; Pred. No. 3.8e-26;

Matches 361; Conservative 213; Mismatches 636; Indels 385; Gaps 60;

Qy 57 PPRIVEHPSDLIVSKGEPATLNCKAECRPTPTIEWYKGERVETDKDDPRSHRMLPSGS 126

Db 21 PLYFLVEPVDTLVSRGSSVILNCAYSSEPSNIEWKKDGTFLNLVSD---RRQLLPDGS 77

Qy 127 LFFLRIVHGRKSRPDEGVYVCVAR-NYLGEAVSHNASLEVAIRLDDFRQNPVDVMVAVGE 185

Db 78 LFTSNVHSHKHKPDEGFYQCVATVNLGTIVSRATAKLAVAL-PRFTSQPESSIVGN 136

Qy 186 PAVMECPQPRGHPEPTISWKKDGLSDDKDERITIRGGKLMITYTRKSDAGKIYCVGTNM 245

Db 137 SGILNCE-VNADLVPPVRWEQNRQPLLLDDRIKVLPSGLTVISNATEGDEGLRCIVESG 195

Qy 246 VGERESVAELTVLERPS-----FVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWKDD 299

Db 196 GPPKFSDEALKVLQSEEMLDLFLMRPSSMIKVIQSAVLPCVAGSLPAPVIRWMKNE 255

Qy 300 GEL---PKSRYEIRDDHTLKIRKVTAGDMGYSYTCVAENMVGKAEASATLVQEPHFVVK 356

Db 256 DVLDTSSGRLLAGGSLEISDVTEDDAGTYFCVADNKNKTIEAQELTVQVPPPEFLKQ 315
Qy 357 PRDQVVALGRVTVFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGLTIT 416
Db 316 PANIYARESDIVFECEVTGKPAPTVMKNGDVVPSDY-----FKIVKEHNLQVL 367
Qy 417 NVQRSDVGYICQTLNAGSIITKAYLEVTDIADIRPPVIRQGPVQTVAVDGTFLVSC 476
Db 368 GLVKSDEGFYQCAIENDVGNAGAGL-----IILEHAP----- 401
Qy 477 VATGSPVPTILWRKDGVLVSTQDSRIKLENGVLQIRYAKLGDTRYTCIASTPSGE-AT 535
Db 402 -ATTGPLPSAPRDVVASLVST-----RFIKL---TWRTASDPHGDNT 441
Qy 536 WSAYIEVQFQGVPPVPPRPDNLIPSAKPEVTVSRNTVTLNWSQPNLNSGATPTSYI 595
Db 442 YSVFYTKE--GVARERVENT-----SQPGEMQVT----- 468
Qy 596 IEAFSHASGSSWQTVANVTETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTO 655
Db 469 -----IQNLMPATVYIFKVMQKNKG-SGESSAPLRVETQ 502
Qy 656 DVLPTSQGVHDKQVQRELGNVHLHNLVTLSSSIEVHW-TVDQSSQYIQGYKILYRPS 714
Db 503 PEV-----QLPGPAPNIRATSTPT-----SITVWETPLSGNGETQNYKLYMEK 548
Qy 715 GANHGSDWLVEFVRTPAKNSVVPDLRKGVNYEIKARPPFNEFGADSEIKFAKTLLEE 774
Db 549 GTDK-EQDQDV-----SSHSYITNGLKYTEYSFRVAYNKHGPGVSTQDVAVRTLSDV 601
Qy 775 PSAPPGQVTVSKNDGNGTAILVSWQPPEDTQNGMVQYKVCNGLNETRYHINKT-VDGS 833
Db 602 PSAAPQNLSEVR--NSKSIVHWQPPSSATQNGITGKYIRKASRKSDVTETVVTGT 659
Qy 834 TFSVVIFPLVPGIRYSVEVAASTGAGSGVKSE-----PQFIQLDAHGNP 877
Db 660 QLSQIEGLDRGTEYNFRVAALTVNGTGPATDWLSAETPESDLDESVPPE-VPSLHVRP 718
Qy 878 V-----SPEDQVSLAQISDVVQKAPFIAGIACAIIIMVFSIWLYRHRKRNGL 928
Db 719 LVTSIVSVWTPPENQ-----NIVVRGAYIGYIGSPHAQTIRVD---YKQR----- 761
Qy 929 TSTYAGIRKV-PSFTFTPTV-TYQRGEAV---SSGGRP-----GLLNISEP-AAQP 974
Db 762 ---YTTIENLDPSSHYVITLKAPNNVGEGLPIESAVTRPHTDTSEVDFVINAPYTPVP 818
Qy 975 WLADTWPTG-----NNHNDSCISCTAGNNGSDSNLTYSRPAACIANYNQDNKQTN 1029
Db 819 DPTPMPPVGVQASILSHDTIRITW-----ADNSLPKHQKITD--SRYTYV--RWKTN 867
Qy 1030 LMLPESTVYGDVLSNKNEMKTFNSPN-----LKDGRFVNPSGQPTPYATTQLIQS 1081
Db 868 --IPANTKYKNAN-ATTLSTLVTLGLKPNLTLYEFVMTVKGRSSTSWMTAHGATFELVPT 924
Qy 1082 NLSNNMNGSGDSGEK---HWKPLGQKQEVAP-VQYNIVEQNKLNDYRANDIVPPTI 1136
Db 925 SPPKDVTVVSKGKPRTIIVNQPPSEANGKITGYIIYSTDVNAEIHWDVIEPVVGNRL 984
Qy 1137 PYN-QSYDQNTGGSYNSSDRGSSTSGSQGHKKGARTPKVPKQGGMNWADLLPPPPA---- 1191
Db 985 THQIQELTLDTPTYYFKIARNKSGMGPMSEAVQFRTPKADS-----SDKMPNDQALGSA 1038
Qy 1192 -----HPPPHSNSEYINISVDESIDQEMPCPVPPA-----RMYLQ 1227
Db 1039 GKGGRLPDLSGDTKPPMSGSNPHGSPTSPLDNMLLVIIISIGVITVVVVIIAIVFCTR 1098
Qy 1228 DELEEEEDERGPTPPVGAASSPAVSYSHQSTAITLTPSQEELQPMQLQDCPEETGHMQH 1287
Db 1099 RTTSHQKKKRAACKSVNG-----SHRYKGNCRDKVPDLWI-----H 1135
Qy 1288 QPDRRRQPVSPPPPRPISPPHTYGISGLVSDMDTDAPEEEDEADME---VAKMQTR 1344

Db 1136 HERLELKPIDKSPDPNPM-----TDTPIPRNSQDITPDVNSMDSNIHQ 1180
Qy 1345 RLLRLGLEQTTPASSVGDLESSVTGSMINGWGSASEEDNISGRSS---VSSSDGSFFTDA 1401
Db 1181 RNSYRGHESEDSMSTLAGRRMRPKMMPFDSQPPQSVRNTPTSTDTMPASSSQCTCDH 1240
Qy 1402 DFAQAVAAAEYAGLKVARRQMDAAGRRHFHASQCRPTSPVSTDSNMSAAVMQKTRPA 1461
Db 1241 QDPEG-ATSSSY-----LASSQEED-----SGQSLPTAHV-----RPS 1272
Qy 1462 KKLKHQPHLRRETYTDDLPPPPVPP---PAIKSPTAQSKTQLE----- 1502
Db 1273 HPLK-----SFAVPAIPPPGPIYDPALPSTPLLSQALNHLHLSVKTASIGTLGR 1323
Qy 1503 ---VRPVPVKLPMSMDARTDRSRDKGSSYKGREV 1534
Db 1324 SRPMPVVPVSAPEVQEAETRMLEDSE-SSYEPDEL 1357

RESULT 7

CAML_RAT

ID CAML_RAT STANDARD; PRT; 1259 AA.
AC Q05695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN LICAM OR CAML1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91372414; PubMed-1894011;
RA Miura M., Kobayashi M., Asou H., Uyemura K.;
RT "Molecular cloning of cDNA encoding the rat neural cell adhesion
RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced
RT by differential splicing.";
RL FEBS Lett. 289:91-95(1991).
CC -!- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC -!- TISSUE SPECIFICITY: THE SHORTER ISOFORM IS PREDOMINANTLY FOUND IN
CC THE BRAIN, WHILE THE LONGER ISOFORM IS FOUND IN THE PERIPHERAL
CC NERVOUS SYSTEM.
CC -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59149; CAA41860.1; -.
DR PIR; S17655; S17655.
DR HSSP; P20241; 1CFB.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
DR PRINTS; PR00014; FNTYPEIII.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Signal; Alternative splicing.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 1259 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1122 EXTRACELLULAR (POTENTIAL).

RA Korenberg J.R., Dreyer W.J.;
RT "Molecular characterization of human neogenin, a DCC-related protein,
RT and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT q23.";
RL Genomics 41:414-421(1997).
CC -1- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -1- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC LINES.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U61262; AAB17263.1;
DR EMBL; U72391; AAC51287.1;
DR MIM; 601907; -.
DR HSPS; P02751; 1TTG.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 1461 NEOGENIN.
FT DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1106 1126 POTENTIAL.
FT DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 67 136 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 166 228 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 263 327 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 355 417 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 436 533 FIBRONECTIN TYPE-III.
FT DOMAIN 536 629 FIBRONECTIN TYPE-III.
FT DOMAIN 630 729 FIBRONECTIN TYPE-III.
FT DOMAIN 735 829 FIBRONECTIN TYPE-III.
FT DOMAIN 850 950 FIBRONECTIN TYPE-III.
FT DOMAIN 951 1052 FIBRONECTIN TYPE-III.
FT DOMAIN 1118 1121 POLY-VAL.
FT DISULFID 74 129 BY SIMILARITY.
FT DISULFID 173 221 BY SIMILARITY.
FT DISULFID 270 320 BY SIMILARITY.
FT DISULFID 362 410 BY SIMILARITY.
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1248 1300 MISSING (IN ISOFORM 2).
FT CONFLICT 168 168 G -> N (IN REF. 2).
SQ SEQUENCE 1461 AA; 159958 MW; 7AAE897E69635A21 CRC64;

Query Match 8.3%; Score 725.5; DB 1; Length 1461;
Best Local Similarity 22.4%; Pred. No. 3.6e-25;

Matches 370; Conservative 208; Mismatches 619; Indels 457; Gaps 58;

Qy 70 IVEHPSDLIVSKGEPATLNCCKAEGRPPTPTIEWYKGGERVETDKDDPRSHRMLPSGSLFF 129
:| | | : :| | | | : | | | | |
Db 56 LVE-PVDTLVSRGSSVILNCSAYSEPSKIEWKDKGTFNLVSDD---RRQLLPDGSLEFI 111
Qy 130 LRIVHGRKSRPDEGVYVCVAR-NYLGEAVSHNASLEVALRDDRFRQNPSPVVMVAVGEPV 188
:| | : :| | | | | | : | | | | | | : |
Db 112 SNVVHSHKHPDEGVYQCVAIVESLGTIIISRTAKLIVAGL-PRFTSQPESSVYAGNGAI 170
Qy 189 MECQPPRGHPPTISWKKDGLDDEIRITIRGKLMITYRKSDAGKYVCVTNMVGE 248
: | : | : :| | | : | : | : | : | : | : |
Db 171 LNCE-VNADLVPPVFWQNRQPLLLDDRVKLPKSGMLVSNATGEGGLYRCVWESGGPP 229
Qy 249 RESEVAELTVLERPS-----FVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDGEL 302
: | : | | | : | : | : | : | : | : | : | : | : |
Db 230 KYSEVELKVLDPDEVISDLVFLKQPSPLVRVIGQDVVLPVAVASGLPTPTIKWKNKEAL 289
Qy 303 ---PKSRYEIRDDHTLKIRKVTAGDMGYSYTCVAENMVGAESATLTVEPPHVVVKPRD 359
: : : :| : | : | : | : | : | : | : | : | : | : | : |
Db 290 DTESSERLVLLAGGSLEISDVTEDDAGTYFCIADNGNETIEAQELTVQAQPEFLKQPTN 349
Qy 360 QVVALGRVTTFQCEATGNPQAFWRREGSQNLFSYQPPQSSSRFSVSQDGLTITNVQ 419
: | : | | | | | : | : | : | : | : | : | : | : | : | : |
Db 350 IYAHESMDIVFECEVTGKPTPTVKKVKNQDMVPSDY-----FKIVKEHNLQVLGLV 401
Qy 420 RSDVGYICQTLNVAGSIITKAYLEVTVDIADRPVPIRQGPVNVAVDGTFLVSCVAT 479
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 402 KSDEGFYQCIENDVGAQAQAQL-----IILEHAP-----AT 434
Qy 480 GSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDTRYTCIASTPSGE-ATWSA 538
: | : : | | | : | : | : | : | : | : | : | : | : | : | : |
Db 435 TGPLPSAPRDVVASLVST-----RFIKL-----TWRTPASDPHGDNLYTSV 475
Qy 539 YIEVQFEGVPVQPPRPTDNPILSAPSKEVTVDSRNTVLSWQPNLNSGATPTSIIIEA 598
: : : : | : : | : | : | : | : | : | : | : | : | : | : |
Db 476 FYTKE--GIARERVENT-----SHPGEMQVT-----499
Qy 599 FSHAGSSWQTVAEVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTDVL 658
: | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 500 -----IQNLMPATVYIFRVMAQNKHG-SGESSAPLRVETQPEV 536
Qy 659 PTSQGVQVQVRELGNVAVLHNPVLSVSSSIEVHW-TVDQSQYIQGYKILYRPSGAN 717
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 537 -----QLPGAPNLRAYASPT-----SITVTWETPVSGNGEIQNKLYMEKGT 582
Qy 718 HGESDWLVFEVTRPAKNSVVPIDLRKGVNIEIKARPPFNEFGADSEIKFAKLEEPSA 777
: | | : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 583 K-EQDVV-----SSHSYTINGLKKYTESFRVYVAYNKHGPGVSTPDVAVRTLSVPSA 635
Qy 778 PPQGVTVSKNDGNGTAILVSWQPPPTDQNGMVQYKVCWCLNETRYHINKT-VDGSTFS 836
: | : : | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 636 APQNLSEVR--NSKSMIHWQPPAPATQNGQITGYKIRYKASRSDVTETLVSGTQLS 693
Qy 837 VVIPFLVPKIRYSVEVAASTAGSGVKSE-----PQFIQLDAHGNPV-- 878
: | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 694 QLIEGLDRGTEYNFRVAALTINGTGPATDWSLAETFSOLDETRVP-VPSSLHVRPLVT 752
Qy 879 -----SPEDQVSLAQISDVVKQPAFIAGIAGAACWILMVFSINLYRHRKRNGLTST 931
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 753 SIVVSWTPPENQ-----NIVRVGAYIGYGISPHAQTIKVD---YKQR----- 792
Qy 932 YAGIRKV-PSFTFTPTV-TYRQGEAV-----SSGGRP-----GLLNISEP-AAQPWLA 977
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 793 YTTIENLDPSHHVITLKAFFNVGEGIPLESYAVTRPHTDTSEVDFVINAPYTPVPDPT 852
Qy 978 DTWPNTG-----NNHNDCSISCTAGNGNSDNLTYTSRPAACIANYNQNDKQNTNML 1032
: | | : : | : : | : | : | : | : | : | : | : | : | : | : | : |
Db 853 PMMPVGVQASILSHDTIRITW-----ADNSLPKHQKIDT--SRYYTV--RWKTN--I 899
Qy 1033 PESTIVYGDVLSN-----KINEMKTFNSPNLKDGRFVNPSPGQTPYATT-QLIQSNL 1093
: | : | : : | : : | : | : | : | : | : | : | : | : | : | : |
Db 900 PANTKYKNANATLSYLVTLGKPNLTLYEFSVMVTKGRR--SSTWSMTAHTGTFLVPTSP 957

Qy 1084 SNNMNGSGDSGEK---HWKPLGQKQEV-----APVQYNIVEQ---NKLNK 1124
Db 958 PKDVTVSKBGKPKTIIVNQPPSEANGKITGYIIYSTDVNAEHDWVIEPVVGNRLT- 1016
Qy 1125 DYRANDTVPTPIPNQSYDQNTGG-----SYNSSDR--GSSTSGSGHKKGA 1169
Db 1017 -HQIQELTLDTPIYFKIQRNSKGMSEAVQFRTPKADSSDKMPNDQASGSG--KGS 1073
Qy 1170 RTPKVPKQGGMNWADLLPPPAHPPPHSNSEEYNSVDESYDQEMPCPVPPA----- 1221
Db 1074 RLPDL-----GSDYKPPMSGSGNSPHG-----SPTSLDSNMLLVIIIVSGVITIV 1119
Qy 1222 ---RMYLQDELEEEEDERGTPPVVGAASSPAVSYSHQSTATLTPSQEELQPLQD 1277
Db 1120 VVIIAVFCTRTTSHQKKRAACKSVNG-----SHKYKGNKSDVKPDLWI----- 1165
Qy 1278 CPEETGHMQHQPDRRRQPVSPPPRPISPPHTYGISGLVSDMDTDAPEEEDEADME 1337
Db 1166 -----HHERLELKPDKSPDPNPM-----TDTPIPRNSQD----- 1196
Qy 1338 VAKMQTRRLRLRGLEQTPASSYGDLESSVTGSMINGWGSASEEDNIS--SGRSSVSSSDG 1395
Db 1197 -----ITPVDNSMDSNHQRRNSYRGH-----ESEDMSMTLAGRGRMRPMM 1238
Qy 1396 SFPTDADFQAQAAAAEYAGLVARRQMDAAGRHHFHASQCPRPTS-----PVS 1445
Db 1239 MPP-DSQPPQPVISAHPHSLD-----NPHHHFSSSLASPARSHLYHGPSWPWIG 1288
Qy 1446 TDSNMS--AAVMQKTRPAKLLHQPHGLRRETYTD-----DL 1480
Db 1289 TMSLSDRANSTESVRNTPSTDTMPASSQTCCTDHQDPEGATSSSYLASSQEEDSGQSL 1348
Qy 1481 P-----PPVPP--PAIKSPTAQSKTQLE----- 1502
Db 1349 PTAHVRPSHPLKSFAPVPAIPGPTTYDPALPSTPLSQALNHHISVKTASIGTLGRS 1408
Qy 1503 --VRPVVVKLPMSDARTDRSSDRKGSSYKGREV 1534
Db 1409 RPPMPVVVPSAPEVQ-ETTRMLEDSSESYEPDEL 1441

RESULT 9

NRCA_CHICK

ID NRCA_CHICK STANDARD; PRT; 1284 AA.
P35331;
01-FEB-1994 (Rel. 28, Created)
01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NG-CAM RELATED CELL ADHESION MOLECULE PRECURSOR (NR-CAM) (BRAVO).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX MEDLINE-91258407; PubMed-2045418;
RA Grumet M., Mauro V., Burgoon M.P., Edelman G.M., Cunningham B.A.;
RT "Structure of a new nervous system glycoprotein, Nr-CAM, and its
RT relationship to subgroups of neural cell adhesion molecules.";
RL J. Cell Biol. 113:1399-1412(1991).
RN [2]
RP SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-EMBRYONIC BRAIN, AND RETINA;
RX MEDLINE-92381110; PubMed-1512296;
RA Kayyem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;
RT "Bravo/Nr-CAM is closely related to the cell adhesion molecules L1
RT and Ng-CAM and has a similar heterodimer structure.";
RL J. Cell Biol. 118:1259-1270(1992).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION; NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC. SPECIFICALLY INVOLVED IN THE DEVELOPMENT OF OPTIC

CC FIBRES IN THE RETINA.
CC -1- SUBUNIT: HETERODIMER, COMPOSED OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RETINA AND DEVELOPING BRAIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN DEVELOPING NEURAL RETINA AND
CC EMBRYONIC BRAIN TISSUE.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58482; CAA41391.1; -.
DR EMBL; L08960; AAA48632.1; -.
DR HSSP; P20241; ICFB.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 5.
DR PFAM; PF00047; ig; 6.
DR PRINTS; PR00014; FNTYPEIII.
KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW Transmembrane; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 1284 NG-CAM RELATED CELL ADHESION MOLECULE.
FT DOMAIN 25 1143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1144 1166 POTENTIAL.
FT DOMAIN 1167 1284 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 56 125 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 155 220 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 261 323 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 351 415 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 445 508 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 599 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 638 699 FIBRONECTIN TYPE-III.
FT DOMAIN 738 799 FIBRONECTIN TYPE-III.
FT DOMAIN 837 906 FIBRONECTIN TYPE-III.
FT DOMAIN 943 1006 FIBRONECTIN TYPE-III.
FT DOMAIN 1057 1114 FIBRONECTIN TYPE-III.
FT DISULFID 63 118 POTENTIAL.
FT DISULFID 162 213 POTENTIAL.
FT DISULFID 268 316 POTENTIAL.
FT DISULFID 358 408 POTENTIAL.
FT DISULFID 452 501 POTENTIAL.
FT DISULFID 543 592 POTENTIAL.
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 995 995 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1059 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 612 621 MISSING (IN ISOFORM AS10).
FT VARSPLIC 1027 1038 MISSING (IN ISOFORM AS12).
FT VARSPLIC 1039 1131 MISSING (IN ISOFORM AS93).
FT VARSPLIC 1202 1205 MISSING (IN ISOFORM AS-CYT2).

```

RESULT 10
DCC_MOUSE
ID DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR.
GN DCC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE=96112625; PubMed=8570174;
RA Cooper H.M., Arnes P., Britto J., Gad J., Wilks A.F.;
RT "Cloning of the mouse homologue of the deleted in colorectal cancer
RT gene (MDCC) and its expression in the developing mouse embryo.";
RL Oncogene 11:2243-2254(1995).
RN [2]
RP REVISIONS.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RA Cooper H.M.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM
CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIRD
CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC AND REMAIN AT THIS LEVEL IN THE ADULT.
CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

Db      389 GVVKSDEGF----- 397
Qy      477 VATGSPVPITILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIASTPSGEATW 536
Db      398 -----YQCVAEENAGNAQS 411
Qy      537 SAYIEVQEFGVPVQPPRPTDP--NLIPSAPSKPEVTDVSRNTVTLSWQPNLNSGATPTS 594
Db      412 SAQLIV-----PKPAIPSSSILPSAPRDVLVPLVSSRFVRLSWRPPAEAKGNIQTF 462
Qy      595 IIEAFSHAGSSSWQTVAENVKTET--SAIKGLKPNAIYFLVLRANAYGISDPSQISDPV 652
Db      463 TV--FFSREGDNRRERALNTTQPGSLQTLVGNLKPMEYTFRVVAYNEWG---PGESSQPI 517
Qy      653 KTQDVLPTSQGVHDKQVRELGNVLAHLHNPTLVSSSSIEVHWTVDDQSQY---IQGYK 708
Db      518 KVA-----TQPELVQPGPVENLHAVST-SPTSILITW---EPPAYANGPVQGYR 562
Qy      709 ILYRPSGANHGESDNLVFEVRTPAKNSVVPDLRKGVNYEIKARPFNEF----- 758
Db      563 L-----FCTEVTSGKEQNIEV---DGLSYKLEGLKFFTEYTLRFLAYNRY 604
Qy      759 -QGADSEIKFAKLTLEEAPSAPPGQVTVSKNDGNGTALLVSWQPPPEDTQNGMVQEYKVMC 817
Db      605 GPGVSTDDITVTLSDVPSAPPQNISLEV--VNSRSIKVSWLPPPSGTQNGFITGYKI-R 661
Qy      818 LGNETRYHINKTVDGSTFSVVIPLVPVGIYRSEVEAASTGAGSGVKSEQPFIQLDAHGNP 877
Db      662 HRKTRRGEMETLEPNNLWYLTFLGLEKGSQYSFQVSAMTVNGTGPSSNWTAE-----TP 716
Qy      878 VSPEDQVSLAQISDVKQP-----AFIAGIAGAACWIILMVFSIW 917
Db      717 ENDLDESQVPDQSSSLHVRPQNCIIMSWTPLPNPNIWVRVGIIGYGVG-----SPYAE 771
Qy      918 LYRHRKRKNGLTSTYAGIRKVPST--FTPTVTYQRGGEAV---SSGRPLGLNISEPAA 972
Db      772 VRVDSQKR-----YYSIERLESSSHYVISLKAFNNAGCVPLYESATTRSITDPTDPVD 825
Qy      973 QPWLADTWPNTG----- 984
Db      826 YPPLLDFFPTSGPDVSTPMLPVGVQAVALTAEAVRSWADNSVPKNQRTSDVRLYTVRW 885
Qy      985 -----NNHNDCSISCCTAG-----NGNSD-----SNLTTY----- 1009
Db      886 RTSFSASAKYKSEDTTSLSYTATGLKPNMTWYEFVMTKNRRSSTWSMTAHATTYEAAPT 945
Qy      1010 -----SRPADCIANYNNQLD-----NKQTNLMLEP-----STVYGD-- 1040
Db      946 SAPKDLTVITREGKPRAVISVWQPPLEANGKITAYILFYTLDKNIPIDDWIMETISGDRL 1005
Qy      1041 ---VDLSKNINEMKTF--NSPNLKDGRFVNPSGQPTPYATTQLIQSNLSNMNNGSGDS 1094
Db      1006 THQIMDLS--LDTMYFRIQARNVKG--VGPLSDPILFRTLKVEHPDKMANQGRHGCG 1060
Qy      1095 GEKHWKLPQQKQEVAPVQYVINEQNKLNDKYRANDTVPTPIPYNSQYDQNTGGSYNSSD 1154
Db      1061 G--YW-----PVDTNLIDRSTLNLP-----PIGQMHPHP--GSVTPQK 1094
Qy      1155 RG-----SSTSGSQGHKKGARTPKVPKQGGMNWADLLPPP 1189
Db      1095 NSNLLVITVVTVGLTVLVVVIVAVICTRRSSAQQRKRATHSGSKRKGSK--DLRPPD 1152
Qy      1190 PAHPPPHNSSEYINVSDESVDQEMPCPVPPARMYLQQDELEEEDERGPTPPVRGAASS 1249
Db      1153 -----LWIIHEEMEMKNIEK-PT-----GTD 1172
Qy      1250 PAAVSYSHQSTATLTSPSQEELQPLQDCPEETGHMQHPDRRRQPVSPPPPPRIPSPH 1309
Db      1173 PAGRDSPIQSQDLTP-----VSHSQSETQMGSKSAS-----H 1205
Qy      1310 TYGYSIGPLVSDMDTDAPEEEDEADMEVAKMQTRRLRLGLEQ-----TPASSVGLD 1362

```


CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A

CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68274; CAA48335.1; -.
DR EMBL; X67734; CAA47963.1; -.
DR PIR; S28830; S28830.
DR MIM; 190197; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
FT Cell adhesion; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1012 AXONIN-1.
FT PROPEP 1013 1040 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 54 118 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 148 216 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 313 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 341 402 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 433 495 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 523 594 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 606 612 GLY/PRO-RICH.
FT DOMAIN 611 706 FIBRONECTIN TYPE-III.
FT DOMAIN 714 809 FIBRONECTIN TYPE-III.
FT DOMAIN 816 908 FIBRONECTIN TYPE-III.
FT DOMAIN 917 1003 FIBRONECTIN TYPE-III.
FT SITE 794 796 CELL ATTACHMENT SITE (BY SIMILARITY).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 1012 1012 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1040 AA; 113393 MW; 254E78DD3C28EFB6 CRC64;

Query Match 7.5%; Score 653.5; DB 1; Length 1040;
Best Local Similarity 24.4%; Pred. No. 3.7e-22;
Matches 227; Conservative 115; Mismatches 400; Indels 187; Gaps 26;

Qy 63 QEDFPPIVHEPDLIV---SKGEPATLNCAEGRPTPTIEWYKGERVETDKDDPRSHR 119
| | | : | : | | | | | : | : | |
Db 32 QTFPGPVFEDQPLSVLPPEESTEEQVLLACRARASPPATYRWKXNGTEM---KLEPGSRH 88
| | | : | : | | | | | : | : | |
Qy 120 MLLPSGSLFFLRIVHGRKSRPDEGVVVCARNYLGEAVSHNASLEVAIRLDDFRQNPDSV 179
| : | : | : | | | | | | | | : | : | : |
Db 89 QLV-GGNL---VIMNPTKAQDAGVTQCLASNPVGTVVSREAILRFGPLQEFSEKEDPV 143
| : | : | : | | | | | | | | : | : | : |
Qy 180 MVAVGEPVMECQPPRGHPPTISWKKDGS---LDDKDERITIRGGKLMITYTRKSDAG 236
| : | | | : | : | : | : | : | : | : | : |
Db 144 KAHEGWGVMLPCNPAPAHYPLGSYRWLLNEFPNFIPTDGRHFVSQTTGNLYIARTNASDLG 203
| : | : | : | | | | | | | | : | : | : |
Qy 237 KYVCVTNMVGERSEV-----AELTVLERPSFVKR-PSNLAVTVDDSAEFKCEA 285
| : | : | : | | | | | | | | : | : | : |
Db 204 NYSCLATSHMDFSTKSVFSKFAQLNLAEDTRLFAPSIRKARPAETALVQGQVTELECPA 263
| : | : | : | | | | | | | | : | : | : |
Qy 286 RGDVPVTVRWKDDGELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGKAEASATL 345
| : | | : | | | | | : | : | : | : | : | : |

Db 264 FGNVPVRIKWRKVDGSL--SPQWTTAEPTLQIPSVSFEDEGTYECEARNKGRDVTQGR 321
Qy 346 TVQEPHPFVVKPRDQVVALGRVTVFCEATGNPQPAIFWRREGSQLNLFYQPPQSSSRF 405
| | | : | : | : | | | : | : | : | : |
Db 322 IVQAQPEWLKVISDTEADIGSNLRWCAAGKPRPTVRLNRG-----EPLASQNRV 373
Qy 406 SVSQTGDLTITNVQRSDVGYICQTLNVAIGSIITKAYLEVTDVIADRPVPPVIRQGPVNQT 465
| | | : | : | : | : | : | : | : | : |
Db 374 EV-LAGDLRFSLKLSLEDGSMYQCAVENKHGTIYASAEALAVQALAPD-----FRLNVPVRL 427
Qy 466 V--AVDGTFFVLSCVATGSPVPTILWRKDGVLSTQDSRIKQLENGVLQIRYAKLGDTRY 523
| : | : | : | : | : | : | : | : | : | : |
Db 428 IPAARGGEILIPCPAPRAKAVVLWSK-GTELLVNSRVTVTPDGTLLIIRNISRSDEGKY 486
Qy 524 TCIASTPSGERTWSAYIEVEFGVPVPPRPTDPNL-----559
| | | : | : | : | : | : | : | : | : |
Db 487 TCFAENFMGKANSTGILSVRDKITLAPSSADINLGNLTQCHASHDPTMDLFTFTWL 546
Qy 560 -----559
Db 547 DDFPIDFDKPGGHYRRTNVKETIGDLTILNAQLRHGGKYTCMAQTVVDSASKEATVLVRG 606
Qy 560 IPSAPSKPEVTDVSRNTVLSWQPNLNSGATPTSIIIEAFSHAGSSWQTAENV-----614
| | | : | : | : | : | : | : | : | : |
Db 607 PPGPPGVVVRDIGDTTILQSWRGFDNHSPIAKYTLQARTPPAG-KWKQVRTNPANIEG 665
Qy 615 KTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTDVLPSTSGVDHKGQVREL 674
| : | : | : | : | : | : | : | : | : | : |
Db 666 NAEATAQVLGLTPWMDYEFVRIASNLGTGEPGSGPKIRTEAAPSVA-----713
Qy 675 NAVLHLHNPVTLSS-----SSIEVHVHTVDQSQSIQ-----GYKILYRPSGANHGESDWL 724
| : | : | : | : | : | : | : | : | : | : |
Db 714 -----PSGLSGGGGAPGELIVNWT--PMSREYQNGDGFYLLSFRQSGSTHWQT---760
Qy 725 VFEVTPAKNSVVI----PDLRGVNYEIKARPFNEFGADSEIKFAKTLSEAPSAPPQ 780
| | | : | : | : | : | : | : | : | : |
Db 761 ---ARVPGADAQYFVYSNESVRPTTPEFVKIRSYNRRGDGPESLTALVYSAEEPRVAPT 817
Qy 781 GV---TVSKNDGNGTAILVSWQPPPEDTQNGMVQYKVV--WCLGNTRYHINKTVDGSTF 835
| | | : | : | : | : | : | : | : | : |
Db 818 KWKAGVSSSEMN-----VTWEPVQDD-MNGILLGYEIRYWKAGDKEAADRVRTAGLDT 871
Qy 836 SVVIFPLVPGIRYSVEVAASTGAGSGVKS 864
| : | : | : | : | : | : | : | : | : | : |
Db 872 SARVSLGHPNTKIHVTVTRAYNRAGTGPAS 900

RESULT 13

AXO1_RAT

ID AXO1_RAT STANDARD; PRT: 1040 AA.
AC P22063;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1).
GN TAX1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC TISSUE=SPINAL CORD;
RX MEDLINE=90199890; PubMed=2317872;
RA Furlay A.J., Morton S.B., Manalo D., Karagogeos D., Dodd J.,
RA Jessell T.M.;
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily
RT member with neurite outgrowth-promoting activity."
RL Cell 61:157-170(1990).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF
CC AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A
CC GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- TISSUE SPECIFICITY: IN NEURAL TISSUES IN EMBRYOS, AND IN ADULT


```
Qy      52 NSLGYTGSLRQLQEDFPPIRVHPESDLIV---SKGEPATLNCKAEGRPTPTIEWYKGGERV 108
       :|::: |||||: |::||:|:| | | | |
Db      23 SSPGSFQAQGTPTATFGPIFEQIGLLFPPEESAEDQVTLACRARASPPATYRWKMNG-- 79

Qy     109 ETDKD-DPRSHRMLLPSGLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAI 167
       ||: |:| |:| :: ||| |:| :| | |
Db      80 -TDMNLEPGSSRHQLM-GGNL----VIMSPTKTQDAGVYQCLASNPGVTGVSKEAVLRFGF 133

Qy     168 LRDDFRQNPSDMVMVAVGEPAMVECPQPRGHPEPTISWKKGDP--LDDKERITIRGGK 224
       |:| :: ||| ::| | :| :| :| ::|
Db     134 LQEFSKEERDPVKTHBGWGMPLPCNPAPHYPGLSYRWLLNEFPNFIPDGRHFVSQTGN 193

Qy     225 LMITYTRKSDAGKYCVGTNMVGERESEV-----AELTVLERPSPVKR-PSNLAV 273
       ||| ||| |:| :| | || | |||
Db     194 LYIARTNASDLGNTSCLATSMDFTSKSVFSKAQLNALAEDPRLPAPSIKARPPETYA 253

Qy     274 TVDSDAEFKCEARGDPVPTVRWKDDGELPKSRYEIRDHTLKIRVKTAGDMGSYTCAE 333
       | :| :| ||| ::||| | | | | | | |
```

```

Db      254 LVGQQVTLECFAGNPVPRIKRWKVDSGL--SPQWATAEPTLQIPSVSFEDEGTYECEAE    311
Qy      334 NMVGKAESATLTVEPPHFVVKPRDQVVALGRVTVFQCEATGNPQPAIFWRREGSQNLL    393
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      312 NSKGRDTVQGRIVQAQPEWLKVISDTEADIGSNLRWGCAAGKPRPMVRWLNRG-----    366
Qy      394 FSYQPPOSSSRFSVSQTGDLTITNVQRSDVGYYICQTLNVAGSIITKAYLEVTDVIADRP    453
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      367 ---EPLASQNRVVE-LAGDLRFSLKSLEDSGMYCVAENKHGTIYASAEVLQAPLAD--    420
Qy      454 PPVIRQGPVNQTV--AVDGTFFLVSCVATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQ    511
       || | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      421 ---FRQPNVRLIPAARGGEISILCPRAAPKATILWSK-GTEILGNSTRVTVTSDGTLI    476
Qy      512 IRYAKLGDTGRYTCTIASTPSGEATWSAYIEVQEFQVGPVQPPRPTDPNL-----    559
       || | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      477 IRNISRSDEGKYTCFAENFMKANSTGILSVRDATKITLAPSSADINVDNLTQLCHASH    536
Qy      560 -----                                559
Db      537 DPTMDLTFTWLDLDDPIDFDKPGGHYRASAKETIGDLTILNAHVHRHGKGYTCMAQTVD    596
Qy      560 -----IPSAPSKPEVTDVSRNTVTLSWQPNLNSGATPTYIIIEAFSHASGSSW    607
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      597 GTSKEATVLVRGGPPGGVVRVIDGTTQLSWSRGFDNHSPIAKYTLQARTPPSG-KW    655
Qy      608 QTVARNV----KTETSAIKGLKPNAIYLFLVRAANAYGISDPQSISDVPKTQDVLPTSQ    662
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      656 KQVRTNPVNIEGNAETAQVLGLMPWMDYEFRRVASNNILGTGEPSPGSSKIRTKEAVPSVA    715
Qy      663 GVDHKQVRELGNVLAHLHNPTVLSS-----SSIEVHWT-VDQOSQYIQGYKILYRPSGA    716
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      716 -----PSGLSGGGGAPGELIINWTPVSVREYNQDGFGF--YLLSFR    753
Qy      717 NHGESDWLVFEVRTPAKNSVVI---PDLRGVNYEIKARPFNEFOGADSEIKFAKLE    772
       | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      754 RQGSSSWQT--ARVPGADAQYFVYGNDISQPYTPFEVKIRSYNRRDGGPESLTALVYSAE    811
Qy      773 EAPSAPPQGVTVSKNDGNGTAILVSWQPPEDTQNGMVQEYKV--WCLGNTRYHINKTV    830
       | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      812 EEPRVAP--AKVWAKGSSSEMNVSWEPVLQD-MNGILLGYEIRYWKAGDNAAAARDVRT    868
Qy      831 DGSTSFSVVIPFLVPIGIRYSVEVAASTGAGSGVKSEPOFIQLDAHGNPVSPEDQVS    885
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      869 AGLDTSARVTLGNPKTYHVTVRAYNRAGTGPA$-PSADAMTVKPPPPRRPGNIS    922

```

AX01 CHICK

```

ID  AXO1_CHECK          STANDARD;          PRT;  1036 AA.
AC  P28685;
DT  01-DEC-1992 (Rel. 24, Created)
DT  01-DEC-1992 (Rel. 24, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  AXONIN-1 PRECURSOR.
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  [1]
RP  SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC  TISSUE=BRAIN;
RX  MEDLINE=92174893; PubMed=1311675;
   Zuellig R.A., Rader C., Schroeder A., Kalousek M.B.,
RA  von Bohlen Und Halbach F., Osterwalder T., Inan C., Stoeckli E.T.,
RA  Affolter H.-U., Fritz A., Hafen E., Sonderegger P.;
RT  "The axonally secreted cell adhesion molecule, axonin-1. Primary
RT  structure, immunoglobulin-like and fibronectin-type-III-like domains
RT  and glycosyl-phosphatidylinositol anchorage.";
RL  Eur. J. Biochem. 204:453-463(1992).
CC  -1- FUNCTION: AXON-ASSOCIATED CELL ADHESION MOLECULE (AXCAM) WHICH
CC  PROMOTES NEURITE OUTGROWTH BY INTERACTION WITH THE AXCAM L1 (G4)
CC  OF NEURITIC MEMBRANE.

```



```
Qy      58 GSRLRQDDPPPRIVEHPS---DLIVSKGE-----PATLNCKAEGRPPTTIEWYKGGVER 108
       || : : ||| : : : : : : : : : : | : |:|:|:| | | | : :
Db      19 GSAESKGNNRRPRITKQPAGGELLFKVAQQNKESDNPFIIIECADQGPEPEYSWIKNGKKF 78

Qy     109 ETDKDDPRSHRMLLPGSLFFLRIVHGKRKSRLPDEGVVVCVARNYLGEAVSHNASLEVAIL 168
       : | | | | | : : | : | | | | | | | : : | : | |
Db      79 DWQAYDNRMLRQ--PGRGTLVTI---PKDERGRHYQCASNEFGTATSNSVIVRKAE 132

Qy     169 RDDFRQNPSDMVMV-GEPAVMECQPPRGHPPTISW----- 204
       : | : : : : | | | : : | | | | : |
Db     133 -NAPKDAAKFLLEAVEGEPMLKCAAPDGFPSPVTWNMMIQESIDGSIKSINNSRMTLDPE 191

Qy     205 -----KKD----- 207
       : |
Db     192 GNLWFSNVTRDEDASSDFYYACSATSVRFSEYKIGNKVLLDVQMGSVASONKHPPVRQYV 251

Qy     208 -----GSPLD----DKD-----ERIT--IRGGKLMITYTRKSD 234
       | : | | | | : | | : | | | : | | | |
Db     252 SRRQSGLALRGKRMELFCIYGGTPLPQTIVMSKDGQRIOQSDRITQGHYKSLVIRQTNFDD 311

Qy     235 AGKYVCVGNTMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVR 294
       || | | : | | | : | | | | | | | | : | : | | | | |
Db     312 AGTYTCDVSNCGNAQSFSIILNVNSVPFTFKEPEIATAADEEVVFECRAAGVPEPKIS 371

Qy     295 WRKDDELGPKSRYEIR---DDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTQV-EP 350
       | : : : | : | : | : | : | : | : | : | | | |
Db     372 WHHGKPIEQSTPNPRRTVDNTIRIINLVKGDGTGNVGCNATNSLGYVKDYVLNWQAEP 431

Qy     351 PHFVVKPRDQVVALGRVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQT 410
       | | | | | | | | | | | | | | | | : | : | : |
Db     432 PTISEAAPAAVSTDGRNVTIKCRVNGSPKPLVKWLRL--ASNWL-----TGGRYNVQAN 482

Qy     411 GDLTITNQRSVDGYIICQLTN-----VAGSIITKAYLEVTDVIADRPPIPVRQGPVNQ 464
       || | | : | | | | | | | | : | : : : | : | | | |
Db     483 GDLEIQDVTFS DAGKYCTYAQNKFGEIQADGSLVVEKHEHTRIT-----QEPONY 530

Qy     465 TVAVDGTfVLSCVATGSPVPTIL--WRKDGVLV--STQDSRIKQLENGVLQIRYAKLGDT 520
       | | | | | | | | | | | | | | : | : : : : : : :
Db     531 EVAAGQSATPCRCNEAHDDTLEIEIDWWKDGQSIDFEAQPRFVKTNDSNLTIAKTMEL-DS 589

Qy     521 GRYTCIASPTSGEATWSAYIEVQEFVGPVQPPRPRTDNPILIPSAPSKEVTDVSRNTVYLS 580
       | | | | : : : : | : | : | : | : | : | : | : : :
Db     590 GEYTCVARTRLDEATEARANLIVQD-----VPNAPKLTGIT-CQADKAEIH 633

Qy     581 WQPNLNSGATPTSIIIEAFSHAGSSSQWTAENV-KTETSAIKGLKPNAIYLFVLRAANA 639
       | : : : : : : : : : : : : | : | : | : | : | |
Db     634 WEQQGDNRSPLHYTHYQFNSTFSPAWDAAEYKVPNTDSSFVVQMSPWANYTFRVIAFNK 693

Qy     640 YGISDPSQISDPVKTDQVLP-----TSOGVD-----HKQV-- 669
       | | | | | | : | : | : | : | : | : | |
Db     694 IGASPPSAHSDSCTQPDVPFKNPDNVVQGTEPNNLVISWTPMPMBIEHNAPNFHYVSW 753

Qy     670 QRELGNVLHLHN----- 682
       : | : | | : |
Db     754 KRDIAPAAWENNFIWRQNNVIADQPTFVKYLIKVVAINDRGESNVAAEEVVGYSGED 813

Qy     683 ----PT-----VLSSSSIEVHWT-VDOQS--QYIQGYKILYRPSGANHGESDWLWFVEV 728
       | | : : | : | : | : | : | : | : | | |
Db     814 RPLDAPINFTRMQITSISTSGYMAWTPVSEESVRGHFKGYKI--QTWTENEAGEGLREIHV 871

Qy     729 RTPAKNSVVI---PDLRKGVNYEIKARPFNEFQMGADSEIKFAKTLEEAPSPAPQGVTVS 785
       | : | / | : | : | : | : | : | : | : | : |
Db     872 KGDTNALVTQFKPDSK---NY-ARILAYNGRFNGPPSAVIDFDIPEGVPS-PVQGL--- 923

Qy     786 KNDG---NGTAILVSWQPP--PEDTQNGMVQEKYVW-----CLGNMETRY--HINKTVD 831
       | : | : | : | : | : | : | : | : | | |
Db     924 --DAYPLGSSAFMLHWKKPLYP---NGKLTGYKIYYEEVKESYVGERREYDPhi---TD 974

Qy     832 GSTFSVVIPLVPGIRYSVEVAASTGAGSG----- 861
       : : : | : | : : | : |
Db     975 PRVTRMKMAGLKPNSKYRISITATTKMGGSEHYIEKTTLKDAVNVPATPSFSWEOLPS 1034
```

Query Match 7.2%; Score 627; DB 1; Length 1239;
Best Local Similarity 22.0%; Pred. No. 7e-21;
Matches 281; Conservative 148; Mismatches 427; Indels 420; Gaps 49;

Qy 862 -----VKSEPQFIQ----- 870
: | | | :
Db 1035 DNGLAKFRINWLPSTEGHPGTHFFTMHRIKGETQWIRENEEKNSDYQEVGGLDPETAYEF 1094
Qy 871 ----LDAHGNPVSPEQVSLAQQISDVVKOPAFIAG--IGAACWII-----LMVFSI 916
: | | | : : : | : | | : : :
Db 1095 RVSVVDGHFNTESATQEID-----TNTVEGPIMVANETVANAGWFIGMMLALAFIILFI 1149
Qy 917 WLYRHRKKRNLSTY-----AGIRKVPS-----FTFTPTVTYQRGGEAVSSGGRPCLLN 966
: | | | : | | | : | | : | | :
Db 1150 IICIIRNRGGKYDVHDRELANGRRDYPEEGPFHEYSQPLDNKSAGRQSVSSANKPGVES 1209
Qy 967 ISEPAAQPWLADTWP 982
: | : | |
Db 1210 DTDSMAEYGDGTGMN 1225

Search completed: January 22, 2001, 12:29:42
Job time: 1283 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:53:03 ; Search time 559.88 Seconds
(without alignments)
345.628 Million cell updates/sec

Title: US-09-540-245A-18
Perfect score: 8724
Sequence: 1 MKWKHVPLVMISLSPN.....VLGGYGERGDNNEELETES 1651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	8724	100.0	1651	4	Q9Y6N7	Q9y6n7 homo sapien
2	8315	95.3	1651	11	O55005	O55005 rattus norv
3	8120	93.1	1612	11	O89026	O89026 mus musculus
4	3104	35.6	1060	11	Q9QZ13	Q9qz13 rattus norv
5	2607.5	29.9	1344	11	Q9Z214	Q9z214 mus musculus
6	1594	18.3	1395	5	Q9W213	Q9w213 drosophila
7	1592	18.2	1395	5	O44924	O44924 drosophila
8	1515	17.4	1273	5	O44928	O44928 caenorhabdi
9	1307.5	15.0	859	5	Q9VP26	Q9vp26 drosophila
10	1212	13.9	823	5	Q9VQ10	Q9vq10 drosophila
11	873	10.0	285	4	O43608	O43608 homo sapien
12	852	9.8	423	5	P91572	P91572 caenorhabdi
13	760	8.7	1493	11	P97798	P97798 mus musculus
14	758	8.7	1259	11	Q9QY38	Q9qy38 mus musculus
15	757.5	8.7	1443	13	Q90610	Q90610 gallus gall
16	756	8.7	1822	4	Q9ULT7	Q9ult7 homo sapien
17	748	8.6	1248	6	Q9XT41	Q9xt41 cercopithec
18	746.5	8.6	1377	11	P97603	P97603 rattus norv
19	726.5	8.3	1427	13	Q91562	Q91562 xenopus lae

20	725.5	8.3	1461	4	Q92859	Q92859 homo sapien
21	725.5	8.3	1461	4	O00340	O00340 homo sapien
22	718.5	8.2	1100	4	O94779	O94779 homo sapien
23	713	8.2	1026	4	O94780	O94780 homo sapien
24	707.5	8.1	1308	4	Q9UHI4	Q9uhi4 homo sapien
25	706.5	8.1	1026	11	Q62845	Q62845 rattus norv
26	704.5	8.1	1180	4	O15051	O15051 homo sapien
27	702.5	8.1	1028	11	Q62682	Q62682 rattus norv
28	702	8.0	1236	4	Q9UHI3	Q9uhi3 homo sapien
29	700.5	8.0	1299	4	Q92823	Q92823 homo sapien
30	696.5	8.0	2221	5	Q9U1M1	Q9u1m1 drosophila
31	695.5	8.0	1028	11	Q07409	Q07409 mus musculus
32	689.5	7.9	1099	11	P97527	P97527 rattus norv
33	689.5	7.9	1166	11	Q9QVN4	Q9qvn4 rattus sp.
34	689.5	7.9	2016	5	Q9NBA1	Q9nba1 drosophila
35	684.5	7.8	874	5	O01632	O01632 caenorhabdi
36	684.5	7.8	2016	5	Q9V4J9	Q9v4j9 drosophila
37	683.5	7.8	1445	11	Q63155	Q63155 rattus norv
38	681	7.8	1299	4	O15179	O15179 homo sapien
39	679.5	7.8	1215	11	P97686	P97686 rattus norv
40	678.5	7.8	2222	5	O97394	O97394 drosophila
41	674	7.7	1154	11	Q9QVN3	Q9qvn3 rattus sp.
42	666	7.6	1028	11	P97528	P97528 rattus norv
43	664.5	7.6	1028	4	Q9Q52	Q9q52 homo sapien
44	657.5	7.5	1224	4	O00533	O00533 homo sapien
45	653	7.5	1028	11	Q9JMB8	Q9jmb8 mus musculus

ALIGNMENTS

RESULT 1
Q9Y6N7
ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6N7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040990; AAC39575.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 100.0%; Score 8724; DB 4; Length 1651;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKWKHVPLVMISLSPNHLFLAQLIPDPEDVERGNHGTPIPTSDNDDNSLGYTGS 60
|||||
Db 1 MKWKHVPLVMISLSPNHLFLAQLIPDPEDVERGNHGTPIPTSDNDDNSLGYTGS 60

Qy 61 LRQEDFPPIVEHPDILVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHRM 120
|||||
Db 61 LRQEDFPPIVEHPDILVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHRM 120

Qy 121 LLPSGSLFFLRIVHGRKSRPDEGVYVYVARNYLGEAVSHNASLEVAILLRDDFRQNPSPDM 180
Db 121 LLPSGSLFFLRIVHGRKSRPDEGVYVYVARNYLGEAVSHNASLEVAILLRDDFRQNPSPDM 180
Qy 181 VAVGEPVMECQPPRGHPEPTISWKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Db 181 VAVGEPVMECQPPRGHPEPTISWKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Qy 241 VGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Db 241 VGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDDG 300
Qy 301 ELKRSRYEIRDDHTLKIRKVTAGDMGYSITCVAENMVGAESATLTVEPPHFVVKPRDQ 360
Db 301 ELKRSRYEIRDDHTLKIRKVTAGDMGYSITCVAENMVGAESATLTVEPPHFVVKPRDQ 360
Qy 361 VVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGLDITNVQR 420
Db 361 VVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGLDITNVQR 420
Qy 421 SDVGYIICQTLNVAGSIITKAYLEVTDVIADRPVPIRQGVNQTVAVDGTFLVSCVATG 480
Db 421 SDVGYIICQTLNVAGSIITKAYLEVTDVIADRPVPIRQGVNQTVAVDGTFLVSCVATG 480
Qy 481 SPYPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGTGRYTCIASTPSGEATSAIY 540
Db 481 SPYPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGTGRYTCIASTPSGEATSAIY 540
Qy 541 EVQEFQVVPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLNQPNLNSGATPTSYIIIEAFS 600
Db 541 EVQEFQVVPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLNQPNLNSGATPTSYIIIEAFS 600
Qy 601 HASGSSWQTVANVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKQDVLP 660
Db 601 HASGSSWQTVANVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKQDVLP 660
Qy 661 SQGVHDKVQRELGNVLAHLNPTVLSSSSIEVHTVDQSQYIQGYKILYRPSGANHGE 720
Db 661 SQGVHDKVQRELGNVLAHLNPTVLSSSSIEVHTVDQSQYIQGYKILYRPSGANHGE 720
Qy 721 SDWLVEFVRTPAKNSVVIDLRKGVNIEIKARPFNEFGADSEIKFAKLEEAPSAPPO 780
Db 721 SDWLVEFVRTPAKNSVVIDLRKGVNIEIKARPFNEFGADSEIKFAKLEEAPSAPPO 780
Qy 781 GVTYSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWLGNETRYHINKTVDSGTFVSVIP 840
Db 781 GVTYSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWLGNETRYHINKTVDSGTFVSVIP 840
Qy 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQFIQLDAHGPNVSPEDQVSLAQQISDVVKQPAFI 900
Db 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQFIQLDAHGPNVSPEDQVSLAQQISDVVKQPAFI 900
Qy 901 AGIGAACWILMVFSIWLYRHRKKRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Db 901 AGIGAACWILMVFSIWLYRHRKKRNGLTSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Qy 961 RPELLNISEPAAQPLADTWPNNGNNDSCISCTAGNGSDSNLTYSRPADCIANYN 1020
Db 961 RPELLNISEPAAQPLADTWPNNGNNDSCISCTAGNGSDSNLTYSRPADCIANYN 1020
Qy 1021 NQLDNKQTNLMLESTVYGDVLSNKNEMKTFNSPNLADGRFVNPSSGQTPYATTQLIQ 1080
Db 1021 NQLDNKQTNLMLESTVYGDVLSNKNEMKTFNSPNLADGRFVNPSSGQTPYATTQLIQ 1080
Qy 1081 SNLSNMNNGSGDSGEKHWKPLGQQKQEVAPVQYNIQVKNLNDYRANDTVPTPIYNY 1140
Db 1081 SNLSNMNNGSGDSGEKHWKPLGQQKQEVAPVQYNIQVKNLNDYRANDTVPTPIYNY 1140
Qy 1141 SYDQNTGGSYNSSDRGSSSTSGSGHKKGARTPKVPKQGGMNWADLLPPPPAHPPPHNSSE 1200
Db 1141 SYDQNTGGSYNSSDRGSSSTSGSGHKKGARTPKVPKQGGMNWADLLPPPPAHPPPHNSSE 1200
Qy 1201 EYNISVDESVDQEMPCVPPARMYLQDELEEEEDERGTPPPVGAASSPAVSYSHQST 1260

Db 1201 EYNISVDESVDQEMPCVPPARMYLQDELEEEEDERGTPPPVGAASSPAVSYSHQST 1260
Qy 1261 ATLTPSQEELQPMQLQDCPEETGHMHQPDRRRQPVSPPPPPRISPPHTYIGYISGPLVS 1320
Db 1261 ATLTPSQEELQPMQLQDCPEETGHMHQPDRRRQPVSPPPPPRISPPHTYIGYISGPLVS 1320
Qy 1321 DMDTAPEEEEDEADMEVAKMQTRRLLRLGLEQTPASSVGDLESSVTGSMINGWSASEE 1380
Db 1321 DMDTAPEEEEDEADMEVAKMQTRRLLRLGLEQTPASSVGDLESSVTGSMINGWSASEE 1380
Qy 1381 DNISGRSSVSSSDGSFFTDADFAQAVAAAEYAGLKVARRQMADAAGRRHFASQCPRP 1440
Db 1381 DNISGRSSVSSSDGSFFTDADFAQAVAAAEYAGLKVARRQMADAAGRRHFASQCPRP 1440
Qy 1441 TSPVSTDSNMSAAVMQKTRPAKLLKHQPGHLRRETYTDDLPPPPVPPPAIKSPTAQSKTQ 1500
Db 1441 TSPVSTDSNMSAAVMQKTRPAKLLKHQPGHLRRETYTDDLPPPPVPPPAIKSPTAQSKTQ 1500
Qy 1501 LEVRPVVVKLPSPMDARTDRSSDRKGSSYKGEVLDRQVDMRTNPGDPREAEQNDG 1560
Db 1501 LEVRPVVVKLPSPMDARTDRSSDRKGSSYKGEVLDRQVDMRTNPGDPREAEQNDG 1560
Qy 1561 KGRGNKAAKRDLPKATHLIQEDILPYCRPTFTPTSNPRDPSSSSSMSRSGSGRQREQA 1620
Db 1561 KGRGNKAAKRDLPKATHLIQEDILPYCRPTFTPTSNPRDPSSSSSMSRSGSGRQREQA 1620
Qy 1621 NVGRRNIAEMVILGGYERGEDNNEELEETES 1651
Db 1621 NVGRRNIAEMVILGGYERGEDNNEELEETES 1651

RESULT 2

O55005

ID O55005 PRELIMINARY; PRT: 1651 AA.

AC O55005;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE TRANSMEMBRANE RECEPTOR ROBO1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SPINAL CORD;

RX MEDLINE=98117249; PubMed=9458045;

RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,

RA Goodman C.S., Tear G.;

RT "Roundabout controls axon crossing of the CNS midline and defines a

RT novel subfamily of evolutionarily conserved guidance receptors.";

RL Cell 92:205-215(1998).

DR EMBL; AF041082; AAC39960.1; -.

DR HSSP; P56276; 1TLK.

DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00041; fn3; 3.

DR PFAM; PF00047; ig; 5.

KW Transmembrane.

SQ SEQUENCE 1651 AA; 180746 MW; FA2452DD46E186B7 CRC64;

Query Match 95.3%; Score 8315; DB 11; Length 1651;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 1560; Conservative 42; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKWKHPFLVMSISLSPNHLFLAQLIPDPEDVERGNDHGTPIPTSDNDNSLGYTGSR 60

Db 1 MKWKHLPLVMSISLTLKSKHLLAQLIPDPEDLERGNDNGTPTSDNDNSLGYTGSR 60

Qy 61 LRQEDFPRIHVPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGERVETDKDDPRSHRM 120

Db 61 LRQEDFPRIIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPRSHRM 120
Qy 121 LLPSGSLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAIRDDFRQNPSDVM 180
Db 121 LLPSGSLFFLRIVHGRKSRPDEGVYICVARNYLGEAVSHNASLEVAIRDDFRQNPSDVM 180
Qy 181 VAVGEPAVMECQPPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Db 181 VAVGEPAVMECQPPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMITYTRKSDAGKYVC 240
Qy 241 VGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVWRKDDG 300
Db 241 VGTNMVGERESKVDVTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTFGWRKDDG 300
Qy 301 ELPKSYEIRDHTLKIRKVTAGDMSYTCVAENMVGAKEASATLTVQEPHFFVVKPRDQ 360
Db 301 ELPKSYEIRDHTLKIRKVTAGDMSYTCVAENMVGAKEASATLTVQEPHFFVVKPRDQ 360
Qy 361 VVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTDGLTITNVQR 420
Db 361 VVALGRTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTDGLTITNVQR 420
Qy 421 SDVGYYICQTLNVAGSIITKAYLEVTDIADRPVPPVIRQGVNQTVAVDGTFLVSCVATG 480
Db 421 SDVGYYICQTLNVAGSIITKAYLEVTDIADRPVPPVIRQGVNQTVAVDGTFLVSCVATG 480
Qy 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDTRYTCIASTPSGEATWSAYI 540
Db 481 SPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGDTRYTCIASTPSGEATWSAYI 540
Qy 541 EVQEFQVFPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSYIIEAFS 600
Db 541 EVQEFQVFPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNSGATPTSYIIEAFS 600
Qy 601 HASGSSWQTVAVENKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTDVLPT 660
Db 601 HASGSSWQTVAVENKTETFAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTDVLPT 660
Qy 661 SQGVHDHQQVRELGNVHLHNPVTLSSSIEVHWTVDDQSQYIQGYKILYRPSGANHGE 720
Db 661 TQGVHDHQQVRELGNVHLHNPVTLSSSIEVHWTVDDQSQYIQGYKILYRPSGASHGE 720
Qy 721 SDWLVEFVRTPAKNSVVIDLRKGVNEYIKARPPFNEFQADSEIKFAKLEEPSAPPQ 780
Db 721 SEWLVEFVRTPTKNSVVIDLRKGVNEYIKARPPFNEFQADSEIKFAKLEEPSAPPQ 780
Qy 781 GVTYSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWCLGNTRYHINKTVDSGTFSSVIP 840
Db 781 SVTYSKNDGNGTAILVSWQPPEDTQNGMVQYKVCWCLGNTRYHINKTVDSGTFSSVIP 840
Qy 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQIQLDAHGNPVSPEQVSLAQGISDVVKQPAFI 900
Db 841 FLVPGIRYSVEVAASTGAGSGVKSEPPQIQLDSHGNPVSPEQVSLAQGISDVVKQPAFI 900
Qy 901 AGIGAACWILMVFSIWLYRHRKKRNLSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Db 901 AGIGAACWILMVFSIWLYRHRKKRNLSTYAGIRKVPSTFTPTVTYQRGGEAVSSGG 960
Qy 961 RPLLNISEPAAQPLADTWPNNGNNDSCISCTAGNGNSDNLTYSRPADCIANYN 1020
Db 961 RPLLNISEPATQPLADTWPNNGNNDSCISCTAGNGNSDNLTYSRPADCIANYN 1020
Qy 1021 NQLDNKQTNLMLPESTVYGDVDSLKNINEMKTFNSPNLKDGRFVNPSSGQPTPYATTQLIQ 1080
Db 1021 NQLDNKQTNLMLPESTVYGDVDSLKNINEMKTFNSPNLKDGRFVNPSSGQPTPYATTQLIQ 1080
Qy 1081 SNLSNMNNGSGDSGEKHKWPLGQQQEVAPVQYINVEQNKLNKYRANDTVPTPIPYNQ 1140
Db 1081 ANLINMNGGDSSEKHKWPPGQQQEVAPVQYINVEQNKLNKYRANDTVPTPIPYNH 1140
Qy 1141 SYDQNTGGSYNSSDRGSSTSGSQGHKGARTPKAPKQGGMNWADLLPPPAHPPPHNSSE 1200
Db 1141 SYDQNTGGSYNSSDRGSSTSGSQGHKGARTPKAPKQGGMNWADLLPPPAHPPPHNSSE 1200

Qy 1201 EYNISVDESVDQEMPCVPVPPARMYLQDDELEEEEDERGPTPPVGAASSPAVSYSHQST 1260
Db 1201 EYSMSVDESVDQEMPCVPVPPARMYLQDDELEEEEAERGPTPPVGAASSPAVSYSHQST 1260
Qy 1261 ATLTPSQEELQPMQLQDCPEETHGMHQHPPDRRQPVSPPPPPRISPPHTYGYISGPLVS 1320
Db 1261 ATLTPSQEELQPMQLQDCPEDLGHMHPDRRQPVSPPPPPRISPPHTYGYISGPLVS 1320
Qy 1321 DMDTDAPEEEEDADMEVAKMQTRLLRLGLEQTPASSVGDESSVTGSMINGWGSASEE 1380
Db 1321 DMDTDAPEEEEDADMEVAKMQTRLLRLGLEQTPASSVGDESSVTGSMINGWGSASEE 1380
Qy 1381 DNISGRSSVSSSDGSFTDADFQAQVAAAAYAGLKVARRMQDAAGRRHFASQCPRP 1440
Db 1381 DNISGRSSVSSSDGSFTDADFQAQVAAAAYAGLKVARRMQDAAGRRHFASQCPRP 1440
Qy 1441 TSPVSTDSNMSAAVMQKTRPAKKLKHQGHLLRETYTDLPPPPVPPPAKSPTAQSKTQ 1500
Db 1441 TSPVSTDSNMSAAVQKARPTKKQKHQGHLLREAYTDLPPPPVPPPAKSPVSKAQ 1500
Qy 1501 LEVRVVPVKLPMSDARTDRSSDRKGSYKGREVLDRQVVDVMTNPGDPREAEQNDG 1560
Db 1501 LEARIMGPKLASIEARADSSDRKGSYKGREALDRQVVDLRTSPGDPREAEQNPNEG 1560
Qy 1561 KGRGNKAAKRDLPKAKTHLIQEDILPYCRPTFTSNPRDPSSSSSMSSRSGSRQREQA 1620
Db 1561 KARGTKAKRDLPKAKTHLIPEDILPYCRPTFTSNPRDPSSSSSMSSRSGSRQREQA 1620
Qy 1621 NVGRRNIAEMQVLGGYERGEDNNEELEETES 1651
Db 1621 NVGRRNMAEMQVLGGFERGEDNNEELEETES 1651

RESULT 3

089026

ID 089026 PRELIMINARY; PRT: 1612 AA.
AC 089026;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DUT1L PROTEIN.
GN ROBO1 OR DUT1L.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Wu M.C., Lowe N., Fordham R., Rabbitts P.;
RT "The mouse homologue of human DUT1L/H-robo1 gene: protein sequence and
chromosomal location."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17793; CAA76850.1; -.
DR HSSP; P56276; iTLK.
DR MGD; MGI:1274781; Robo1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1612 AA; 176406 MW; 5F2988C544796B4B CRC64;

Query Match 93.1%; Score 8120; DB 11; Length 1612;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1525; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

Qy 57 TGSRLRQEDFPRIIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPR 116
Db 18 SGRSLRQEDFPRIIVEHPSDLIVSKGEPATLNCKAEGRPPTTIEWYKGGERVETDKDDPR 77
Qy 117 SHRLLPSGSLFFLRIVHGRKSRPDEGVYVCVARNYLGEAVSHNASLEVAIRDDFRQNP 176

Db 78 SHRMLLPSGSLFFLRIVHGRKSRPDEGVYICVARNYLGEAVSHNASLEVALRDDFRQNP 137

Qy 177 SDVMVAVGEPAYMECQPPRGHPEPTISWKDGSPLDDKDERITIRGGKLMITYTRKSDAG 236

Db 138 SDVMVAVGEPAYMECQPPRGHPEPTISWKDGSPLDDKDERITIRGGKLMITYTRKSDAG 197

Qy 237 KYVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWR 296

Db 198 KYVCVGTNMVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWR 257

Qy 297 KDDGELPKSRYEIRDHTLKRKVTAGDMGSYTCVAENMVGAASATLTQEPPEFVVK 356

Db 258 KDDGELPKSRYEIRDHTLKRKVTAGDMGSYTCVAENMVGAASATLTQEPPEFVVK 317

Qy 357 PRDQVVALGRTVTFQCEATGNPQPAIFWRREGSNLLFSYQPPQSSSRFSVSQDGLTIT 416

Db 318 PRDQVVALGRTVTFQCEATGNPQPAIFWRREGSNLLFSYQPPQSSSRFSVSQDGLTIT 377

Qy 417 NVQSDVGYICQTLNVAGSIITKAYLEVDVIADPPVIRQGPVNTVAVDQTFVLSC 476

Db 378 NVQSDVGYICQTLNVAGSIITKAYLEVDVIADPPVIRQGPVNTVAVDQTLILSC 437

Qy 477 VATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGTGRYTCTASTPSGEATW 536

Db 438 VATGSPVPTILWRKDGVLVSTQDSRIKQLENGVLQIRYAKLGTGRYTCTASTPSGEATW 497

Qy 537 SAYIEVQFEGVPVQPPRPTDNLIPSAPSKEVTDVSRNTTILSWQPNLNSGATPTSII 596

Db 498 SAYIEVQFEGVPVQPPRPTDNLIPSAPSKEVTDVSRNTTILSWQPNLNSGATPTSII 557

Qy 597 EAFSHASGSSWQTAENVKETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTD 656

Db 558 EAFSHASGSSWQTAENVKETSAIKGLKPNAILFLVRAANAYGISDPSQISDPVKTD 617

Qy 657 VLPTSQGVQVRELGNVHLHNPVTLSSSSIEVHWTVQDQSQYIQYKILYRPSGA 716

Db 618 VLPTSQGVQVRELGNVHLHNPVTLSSSSIEVHWTVQDQSQYIQYKILYRPSGA 677

Qy 717 NHGESDVLVFEVTRPAKNSVYIPDLRKGNYEIKARPPFFNEFGADSEIKFAKLEEAPS 776

Db 678 NHGESDVLVFEVTRPAKNSVYIPDLRKGNYEIKARPPFFNEFGADSEIKFAKLEEAPS 737

Qy 777 APPQGVTVSKNDGNGTAILVSTQPPEDTQNGMVQYKVMWCLGNTRYHINKTVDGSTFS 836

Db 738 APPQGVTVSKNDGNGTAILVSTQPPEDTQNGMVQYKVMWCLGNTRYHINKTVDGSTFS 797

Qy 837 VVIFPLVPGIRYSVEVAASTGAGSGVKSEPFQIQLDAHGNVPSPEQVSLAQQISDVVK 896

Db 896 VVIFPLVPGIRYSVEVAASTGAGSGVKSEPFQIQLDAHGNVPSPEQVSLAQQISDVVK 857

Qy 897 PAFIAGIGAACWILMVFSIWLIRHRRKRNGLSTYAGIRKVPSTFTPTVTYQRGGEAV 956

Db 858 PAFIAGIGAACWILMVFSIWLIRHRRKRNGLSTYAGIRKVPSTFTPTVTYQRGGEAV 917

Qy 957 SSGRPGLLNISEPAAQPLADTWPNNGNHNDCSISCTAGNGNSDNLTTYSRPADCI 1016

Db 918 SSGRPGLLNISEPAAQPLADTWPNNGNHNDCSISCTAGNGNSDNLTTYSRPADCI 977

Qy 1017 ANYNNQLDNKQTNLMLPESTVYGDVLSNKNEMKTFNSPNLKDGRFVNPSPQPTPYATT 1076

Db 1037 ANYNNQLDNKQTNLMLPESTVYGDVLSNKNEMKTFNSPNLKDGRFVNPSPQPTPYATT 1037

Qy 1077 QLIQSNLSNMNNGSGDSSEKHWKPLGQKQEVAPVQYINVEQNLKNKYRANDTVPPTI 1136

Db 1038 QLIQSNLSNMNNGSGDSSEKHWKPLGQKQEVAPVQYINVEQNLKNKYRANDTVPPTI 1097

Qy 1137 PYNQSYDQNTGGSYNSDRGSSTSGSQGHKGARTPKVPKGGGMWADLLPPPAHPPPH 1196

Db 1098 PYNQSYDQNTGGSYNSDRGSSTSGSQGHKGARTPKVPKGGGMWADLLPPPAHPPPH 1157

Qy 1197 SNSEYINISVDESVDQEMPCVPPAPMYLQDELEEEEDERGTPPVVGAASSPAVSY 1256

Db 1158 SNSEYINISVDESVDQEMPCVPPAPMYLQDELEEEEDERGTPPVVGAASSPAVSY 1217

Qy 1257 HQSTATLTPSPQEEELQMLQDCPEETGHMQHQPDRRRQPVSPPPPPRISPHTYGYISG 1316

Db 1218 HQSTATLTPSPQEEELQMLQDCPEEDLGMPHPPDRRRQPVSPPPPPRISPHTYGYISG 1277

Qy 1317 PLVSDMDTDAPEEEEDADMEVAKMQTRLLRLGLEQTPASSVGDLESSVTGSMINGWS 1376

Db 1278 PLVSDMDTDAPEEEEDADMEVAKMQTRLLRLGLEQTPASSVGDLESSVTGSMINGWS 1337

Qy 1377 ASEEDNISSGRSSVSSDGSFFTDADFAQAVAAAAEYAGLKVARRQMDAAGRHHFASQ 1436

Db 1338 ASEEDNISSGRSSVSSDGSFFTDADFAQAVAAAAEYAGLKVARRQMDAAGRHHFASQ 1397

Qy 1437 CPRPTSPVSTDSNMSAAVMQKTRPAKKLKHQPHLRRETYTDDLPPPPVPPPAKSPTAQ 1496

Db 1398 CPRPTSPVSTDSNMSAAVMQKTRPAKKLKHQPHLRRETYTDDLPPPPVPPPAKSPTAQ 1457

Qy 1497 SKTQLEVRPVVVKLPSMDARTDRSSDRKGSYKGREVLDRQVVDVMTNPGDPREAEQ 1556

Db 1458 SKAQLVRPVVVKLASIEARTDRSSDRKGSYKGREVLDRQVVDVMTNPGDPREAEQ 1517

Qy 1557 QNDGKGRGNKAAKRDLPKAKTHLIQEDILPYCRPTFTSNPRDPSSSSSSSRSGSRQ 1616

Db 1518 QNDGKGRGNKAAKRDLPKAKTHLIQEDILPYCRPTFTSNPRDPSSSSSSSRSGSRQ 1577

Qy 1617 REQANVGRNNAEMQVLGGYERGEDNNEELETES 1651

Db 1578 REQANVGRNNAEMQVLGGYERGEDNNEELETES 1612

RESULT 4

Q9QZ13

ID Q9QZ13 PRELIMINARY; PRT; 1060 AA.

AC Q9QZ13;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID:10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-99200391; PubMed-10102268;

RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,

RA Tessier-Lavigne M., Kidd T.;

RT "Slit proteins bind Robo receptors and have an evolutionarily

RT conserved role in repulsive axon guidance.";

RL Cell 96:795-806(1999).

DR EMBL; AF182037; AAF04558.1; -.

DR HSSP; P56276; 1TLK.

DR INTERPRO; IPR001547; -.

DR INTERPRO; IPR001777; -.

DR INTERPRO; IPR003006; -.

DR PFAM; PF00041; fn3; 3.

DR PFAM; PF00047; ig; 5.

DR PRINTS; PR00014; FNTYPEIII.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.

KW Receptor.

FT NON_TER 1060 1060

SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C11E8542DA4 CRC64;

Query Match 35.6%; Score 3104; DB 11; Length 1060;
Best Local Similarity 56.6%; Pred. No. 2.4e-199;
Matches 613; Conservative 150; Mismatches 252; Indels 68; Gaps 19;

Qy 58 GSRLRQEDFPFPRIVEHPSDLIVSKGEPATLNCKAEGRPPTI-----EWYKGGERV 108

Db 21 GSRLRQEDFPFPRIVEHPSDLIVSKGEPATLNCKAEGRPPTI-----EWYKGGERV 75


```

RESULT 5
Q9Z2I4
ID Q9Z2I4 PRELIMINARY; PRT; 1344 AA.
AC Q9Z2I4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE RIG-1 PROTEIN.
GN RBIG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF060570; AAD11628.1; -.
DR HSSP; P56276; 1TLK.
DR MGD; MGI:1343102; Rbig1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1344:AA; 143439 MW; 8B0060341C49CFEA CRC64;

Query Match 29.9%; Score 2607.5; DB 11; Length 1344;
Best Local Similarity 38.2%; Pred. No. 6.5e-166;
Matches 594; Conservative 204; Mismatches 456; Indels 299; Gaps 31;

Qy 58 GSRLRQEDFPFPRIVEHPSDLIVSKGEPATLNCKAAGRPTTIEWYKGGERVETDKDDPRS 117
Db 32 GSRVGPEDAMPRIVEQPPDLVVSRGEPATLPCRAEGRPRPNIEWYKNGARVATAREDPRA 91

Qy 118 HRMLLPSGSLFFLRIVHGRKSRPDGEVYVCARNYLGEAVSHNASLEVAAILRDDFRQNP 177
Db 92 HRLLLPSGALFFPRIVHGRSRPDGEVYTCARNYLGAASRNASLEVAIVLRDDFRQSPG 151

Qy 178 DVMVAVGEPAVMECPQPRGHPEPTISWKKDGSPLDDKDERITIRGGKLMTITYRKSADAG 237
Db 152 NVVVAVGEPAVMECPVPKGHPPLVTKWKGKIKLEEGRITIRGGKLMSHTFKSDAGM 211

Qy 238 YVCVGTNMVGERESEVAELTVLERPSFVKRPSLAVTVDDSAEFKCEARGDPVPTVRWK 297
Db 212 YNCVASNMAGERESGAELVLERPSFLRRPINQVVLADAPVNFLECEVQDQPQNLHWRK 271

Qy 298 DDGELPKSRYEIRDDHTLKIRKVTAGDMGSYTCVAENMVGKAESATLTQVEPPHFVVKP 357
Db 272 DDGELPAGRYEIRSDHSLWIDQVSSDEGTYTCVAENSVGRAESGSLSVHVPQFVTKP 331

Qy 358 RDQVVALGRVTVFQCEATGNQPAIFWRREGSNQLLSYQPPQSSSRFSVSQTDGLTITN 417
Db 332 QNQTVAPGANVSFQCEKGNPPAIFWQKEGSQVLLPSPQSLQPMGRLLVSPRGQLNITE 391

Qy 418 VQRSDVGYYICQTLNVAGSIITKAYLEVTVDIADRPVPIVQGPVNQTVAVDGTFLVSCV 477
Db 392 VKIGDGGYYVCQAVSVAGSILAKALLEIKGASIDGLPIILQGPANQTLVLGSSVWLPCR 451

Qy 478 ATGSPVPTILWRKDGVLVTQDSRIKQLENGVLQIRYAKLGTGRYTCIASTPSGEATWS 537
Db 452 VIGNPQPNIQKKDERWLQGDSDQFNLMNGTLHIASIQEMDMGFYSCVAKSSIGEATWN 511

Qy 538 AIYIEQVE-FGVVPVQPPRPDTPNLPISAPSKPEVTVDSRNTVLSWQPNLNSGATPTSII 596
Db 512 SWLRKQEDWG--ASPGPATGPSNPPGPSQPIVTEVTANSITLTKWPNPQSGATATSYVI 569

Qy 597 EAFSHASGSSWQTVANENKTTETSAIKGLKPNAIYFLVRAANAYGISDPSQISDPVKTD 656
Db 570 EAFSQAAGNTRIVADSGVLETTYISGLQPNITYFLVRAVAGANGLSEPSVPSEVPQTD 629

Qy 657 VLPTSQGVHDHKVQORELGNVHLHNPVLSSSSIEVHWTVDDQSQSIQGYKILYRPSGA 716

```



```

      | : | : || : | : || | | |
Db 1177 -----SNPLSAVAGGTQNR YQITPTNQH--PPQLPAYFATTG 1211

      | : | : || : | : || | | |
Qy 1284 -----HMQHQPDRRR-----QPVSPPPP-PRP-- 1304

      | : | : || : | : || | | |
Db 1212 PGGAVPPNHLFFATQRHAASEYQAGLNAARCAQSACNSCDALATPSPMQPPPPVPVPEG 1271

      | : | : || : | : || | | |
Qy 1305 -ISPPHTYGYISGPLVSD 1321

      | : | : || : | : || | | |
Db 1272 WYQPVHPNSHPMHTSSN 1289

```

044924

```

ID O44924 PRELIMINARY; PRT; 1395 AA.
AC O44924;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.;"
RL Cell 92:205-215(1998).
DR EMBL; AF040989; AAC38849.1; -.
DR HSSP; P56276; ITLK.
DR FLYBASE; FBgn0005631; robo.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEBIII.
SO SEQUENCE 1395 AA; 151778 MW; B820E234A5218983 CRC64;

```

Query Match 18.2%; Score 1592; DB 5; Length 1395;
Best Local Similarity 30.0%; Pred. No. 7.6e-98;
Matches 420; Conservative 187; Mismatches 483; Indels 308; Gaps 40;

```
Qy      68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTPIEWYKGGERVETDKDPRSHRMLPSSGL 127  
       |||:|||:||| | ||||| ||:| |||||:| || | :|:| |:|  
Db      56 PRIIEHPDTLVVKNPEPATLNCKVEGKPEPTIEWFKDGPEPVST--NEKKSHRVQFKDGA 113  
  
Qy     128 FFLRIHVGRKRSPDEGVTVCVARNYLGEAVSHNASLEVAILLRDDFRQNPSDMVMVAVGEPA 187  
       || | :| | :| | |||:| :||| :|||::|:|||| | | || |  
Db     114 FFYRTMQKKQKQ-DGGEYTCVAKNRVQGAVSRHASLQI AVL RDDFRVP E PK DTR VAKGETA 172  
  
Qy     188 VMECQPGRGHPEPTISWKKGDSLDD-----KDERITI-RGGKLMTITYRKSDAGKYV 239  
       ::|| |||:||||:| || |||| | :| | || |:| | || |  
Db     173 LLECQPPKGIPEPTLIWLWIDGVLDDLKAMSGASSRVIRVDGNNLLISNVEPIDEGNYK 232  
  
Qy     240 CVGTNMVGERSSEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRKDD 299  
       |:|:| ||| |:| |:| |:| |:| :| :| :| |:| |:|:|:  
Db     233 CIAQLNVGTRESSYAKLIVQVKPYFMKEPKDQMVLVQTATFHCSVGGDPPPKVLWKKEE 292  
  
Qy     300 GELPKSRYEI-RDDHTLKIRKVTAGDMGSYTCAENMVGKAESATLTVQEPPHFVVKPR 358  
       |:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|:|:  
Db     293 GNIPVSRARILHEKSLAISNITPTDEGYTCEAHNNVGQISARASLIVHAPPNFTKRPS 352  
  
Qy     359 DQVVALGRITVFQCEATGNPQAIFWRREGSQNLLFSYQPQSSRSFVSQTDGLTITNV 418  
       |:|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|:|:  
Db     353 NKKVGLNGVVOLPCMASGNPPPSFVTWTKGVSTLMF---PNSSHGROYTVAAADGTLOITDV 409
```



```

RESULT 9.
Q9VPZ6
ID Q9VPZ6 PRELIMINARY; PRT; 859 AA.
AC Q9VPZ6;
  01-MAY-2000 (TrEMBLrel. 13, Created)
  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG5423 PROTEIN (FRAGMENT).
GN CG5423.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Euphydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bengtson P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

Query Match 15.0%; Score 1307.5; DB 5; Length 859;
Best Local Similarity 33.9%; Pred. No. 4.2e-79;
Matches 300; Conservative 152; Mismatches 345; Indels 87; Gaps 22;

```

Qy 68 PRIVEHPSDLIVSKGEPATLNCKAEGRPPTPTTIEWYKGGERVETDKDDPRSHRMLPSSGL 127
| | | | | | | | | | : | | | | | | | | | | | | | : | | | | | | | | |
Db 1 PRIVEHPIDITTVPRHEPATLNCKAEGSPPTPTTIQWKDGVPL---KILPGSHRITLPAAGGL 57

Qy 128 FFLRIVHGKRKSRPDEGVTVCVARNYLGEAVSHNASLEVAILLRDDFRQNPSDVMVAVGEPA 187
| | | : | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 FFLKVSDDR-----CAVLRDEFRLEPQNTRIAQGDTA 90

Qy 188 VMECQPPRGHPPTTISWKKDGSPLD---DKDERITIRGGKLMITYTRKSDAGKYVCVGTN 244
: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91 LLECAAPRGITPPTVTKWKGQKLDLEGGSKRVRI-VDGNNLAIQDARQTEBGYQCIANK 149

Qy 245 MVGERESEVAELTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVRWRK---DDGEL 302
| | | | | : | | | | : | | : | | : | | | | | | | | | | | : |
Db 150 PVGVRESSLATLVKHVKPYIIRGPHDQTVLEGASVTFPCRVGDDPMPDLWLRLTASGGNM 209

Qy 303 PKSYREIRDDHILKIRKVTAGDMSGYTCAENVMVGKAEASATLTQVEPPHFVVKPRDQVV 362
| | | : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db 210 PLDRVSVLEDRSLRLERVTIADGEYSCEADNVGAITAMGTLTVYAPPKFIQRPAKSKV 269

Qy 363 ALGRVTVTFQCEATGNPQPAIFWRREGSQNLFSYQPPQSSSRFSVSQTGD---LTITNV 418
| | | : | : | | | | | | : | | | | | | | | | | | : | | : |
Db 270 ELGADTSFECRAIGNPKPTIFWTIKNNSTLIFPGAPP--LDRFHSLNTEEGHSILTLTRF 327

Qy 419 QRSQDGYGII-COTLNAGVISIITKAYLEVIDVIADRPVPPVIRQGPVMTAVDGTFFVLSCV 477
| | : | | | : | | : | | | | | | | | | | | | | | | | | |
Db 328 QRTDKDLVILCNAMNEVASITSRVQLSL-DSQEDRPPPIIISGPMVMTLPIKSLATLQCK 386

Qy 478 ATGSPVPTILWRKDGVLSTQDSRIKQLENGVLQIR-YAKLGDITGRYTCIASTPSGEATW 536
| | | | | | | : | | | | : | | | | | | | | | | | | | | | |
Db 387 AIGLPSPTISWYRDGPIQ-PSSKLNITTSGLDIISDLRQQDQGLYTCVASSRAGKSTW 445

Qy 537 SAYIEVQEFQGVQVQPPRPRTDPNL-----IPSAPSKPEVIDVSRNTVTLWSOPNLN 586

```


Db 643 DSTSMKLTWQVNCRLTDGSAIAPHSIAHRHLIRSASFLMQIINGKYVEGYFYARQLPNP 702
Qy 713 -----PSGANHGESDWLVEFVRTP- 731
Db 703 IVNPNAPVTSNTNPLLGSTSTASASASALISTKPNIAAGKRDGETNQSOGGAPTPL 762
Qy 732 -----AKNSVPIPLDLRKGVNYEIKARPFNEFQADSEIKFAKTEE 773
Db 763 NTKYRMLTILNGGASSCTITGLVQTYLIEFFIVPPYKSVGKPSNSRIARTLED 817

RESULT 11
O43608
ID O43608 PRELIMINARY; PRT; 285 AA.
AC O43608;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
T 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
E 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
E ROUNDABOUT 2 (FRAGMENT).
GN ROBO2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors."
RL Cell 92:205-215(1998).
DR EMBL; AF040991; AAC39576.1; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 1.
DR PFAM; PF00047; ig; 2.
FT NON_TER 1 1
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 30606 MW; 05DF916A3DBA96C6 CRC64;

Query Match 10.0%; Score 873; DB 4; Length 285;
Best Local Similarity 59.1%; Pred. No. 1e-50;
Matches 165; Conservative 40; Mismatches 72; Indels 2; Gaps 1;

360 QVVALGRTVTFQCEATGNPQPAIFWRREGSQNLLFSYQPPQSSSRFSVSQTGDLTITNVQ 419
Db 1 QIVAQGRVTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQNSRCSVSPTGDLTITNIQ 60
Qy 420 RSDVGYYICQLTNVAGSIITKAYLEVTDVIAIRPPVIRQGVNQTVAVDGTFLVCVAT 479
Db 61 RSDAGYYICQALTAVGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
Qy 480 GSPVPTILWRKDCVLVSTQDSRIKQLENGVLQIRYAKLGDGTGRYTCIASTPSGEATWSAY 539
Db 121 GDPILPVISWLKEGFTFPGDRPRATIQEQGTILQIKNLRIISDTGTTCVATSSSGEASWSAV 180
Qy 540 IEVQEGVPVQPPRPTDNPILPSAPSKEPTVDSRNTVTLSWQPNLNGATPTSYYIEAF 599
Db 181 LDVTESGATIS--KNYDLSLPGPPSKQVTDVTKNSVTLWSQPGTPTLPASAYIEAF 238
Qy 600 SHASGSSWQTVAENVKTETSIAIKLKPNAIYFLVRAAN 638
Db 239 SQSVNSWQTVANHVKTLYTVRGLRPNTIYLFMVRAIN 277

RESULT 12
P91572
ID P91572 PRELIMINARY; PRT; 423 AA.
AC P91572;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)

DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE SIMILAR TO THE IMMUNOGLOBULIN SUPERFAMILY.
GN ZK377.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nhan M., Hawkins J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88183; AAB52658.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00047; ig; 4.
SQ SEQUENCE 423 AA; 46544 MW; EB4530DB6BD575E5 CRC64;

Query Match 9.8%; Score 852; DB 5; Length 423;
Best Local Similarity 46.1%; Pred. No. 4.7e-49;
Matches 177; Conservative 55; Mismatches 144; Indels 8; Gaps 7;

Qy 68 PRIVEHPSDLIVSKGEPATLNCAGRPTPTIEWYKGGERVETDKDDPSHRMLLPSSGL 127
Db 30 PVIEHPIDVVSRRGSPATLNCAGK-PSTAKITWYKDGQPVITNKQVNSHRIVLDTGSL 88
Qy 128 FFLRIVHGRKSR-PDEGVYVCVARNYLGEAVSHNASLEVAIRLDDFRQNPDSVMVAVGEP 186
Db 89 FLKLVNSGKNGKSDAGAYTCVASNEHEVKSNEGSLKLAIRLDDFRVRPRTVQALGEM 148
Qy 187 AVMEQPPRGHPEPTISWKKDGPLDDK-ERITIRG-GKLMITYTRKSDAGKYVCVGTN 244
Db 149 AVLECSPPRGPEPVVSWRDKKELRIQDMPTTLHSDGNLIIDPVRSDSGTYQCVANN 208
Qy 245 MVGERESEVABLTVLERPSFVKRPSNLAVTVDDSAEFKCEARGDPVPTVWRKDDGELPK 304
Db 209 MVGERVSNPARLSVFEPKFEQPKMDTVDGAVALFDCRVTGDPQPIQTKRKNEMPV 268
Qy 305 SR-YEIRDDHILKIRKVTAGDMGSSYTCVAENMVGAESATLTVQEPHFVVKPRDQVVA 363
Db 269 TRAYIAKNRGLRIERVQPSDEGEYVCYARNPAGTLEASAHRLVQAPPSTQTKPADQVSP 328
Qy 364 LGRTVTFQCEATGNPQPAIFWRREGSQNLLF-SYQPPQSSSRFSVSQTGDLTITNVQRSD 422

```

Qy 170 IVEHPSDLIVSKGEPATLNCCKAEGRPTPTIEWYKGGERVETDKDDPRSHRLMLPGSGLFF 129
   :||| |: |: ||| |:| ||| |: |: || |||| |||
Db 167 LVE-PVDTLVSRGSSVILNCASAYSESPNIEWKKGDTFLNLESD---RQLLPDGSGLFI 122

Qy 130 LRIHVGRKSRPDEGVYVCVAR-NYLGEAVSHNASLEVAIRDDFRQNPSDMVMVAVGEPAV 188
   :|| :||| ||| |: || || ||| | | || |:
Db 123 SVMVSHKHNKPDEGFTYQCVATVDNLGTIVSRTAKLTIVAGL-PRFTSQPEPSSVYVGNASAI 181

Qy 189 MECQPPRGHPPTISWKKDGSLDDKDERITIRGGKLMITYTRKSDAGKYVCVGTNMVGE 248
   :|: |:|: || |: |:| |:| |:| |:| :
Db 182 LNCE-VNADLVPFVRWEQNRQPLLLDDRIIVKLPSGTLVISNATEGDGLLYRCIVESGGPP 240

Qy 249 RESEVAELTVLERPS-----FVKRPSNLAVTDDSAEFKCEARGDPVPTVVRWKDDGEL 302
   :|: ||| ||: | |:|: || |' ||| ||| |:|
Db 241 KFSDEAEKLVLDQPEEIVDLVFLMRPSSMMKVTSQASVLPCVVSGLPAPVVRWMKNEEVL 300

Qy 303 ---PKSRYEIRDHHTLKIRKVTAGDMSGYTCAENMVGKAEASATLTQPEPPHFVVKPRD 359
   | : |:| || |:| |:| || ||| || |:| :
Db 301 DTSSSRLVLLAGGCLEISDVTEDDAGTYFCIADNGNKTVEAQAEALTQVPPGFLKQPAN 360

Qy 360 QVVALGRVTFTQCEATGNPQPAIFWRREGSQNLFSYQPQSSSRFSVSQTGDLTITNVQ 419
   :|:| ||| |:| :| : | | : : |:| :
Db 361 IYAHESMDIVFECEVTGKPTPTVKVKNGDVVI-----PSDNFKVIRKEHLQVLGLV 412

Qy 420 RSDVGYYICTLNVAGSIITKAYLEVT--DVIARDPPPVIRQGPVNQTVAVDGTFLVSCV 477
   :|| |:| |:| |:| : || || | |
Db 413 KSDEGFYQCIAENDVMAQAQAQLIILEHDVAIPTLPP-----TSLISATTDHLAP 463

Qy 478 ATGSPVPTILWRKDCVLVSTQDSRIKOLENGVLQIRYAKLGDTGRYTICIASTPSGE-ATW 536

```

Db	464	ATTGPLSPAPRDVVASLVST-----RFIKL---TWRTPASDPHGDNLT	504
Qy	537	SAYIEVQEFQVGPVQPPRPTDNLIPSAPSKPEVTDVSRNTVTLWSQPNLNGATPTSYII	596
Db	505	SVFYTKE--GVDREVENT-----SQPGEMQVT-----	530
Qy	597	EAFSHASGSSWQTVAEVNKTETSAIKGLKPNAIYLFVRAANAYGISDPSQISDPVKTD	656
Db	531	-----IQNLMPATVYIYFKVMAQNHKG--SGESSAPLRVETQP	565
Qy	657	VLPTSGVDHKQVQRELGNVLHLNPTVLSSSSIEVHW--TVDQSQSIQGYKILYRPSG	715
Db	566	EV-----QLPGPAPNIRAYATSPT-----SITVTWETPLSGNGEIQNYKLYMEKG	611
Qy	716	ANHGESDNLVFEVRTPAKNSVVIDPLRKGVNYEIKARPFNEFGADSEIKFAKTLLEAP	775
Db	612	TDK--EQDIDV-----SSHSTYINGLKKYTEYFVRVAYNKHGPGVSTQDVAVRTLSDVP	664
Qy	776	SAPPQGVTVSKNDGNGTAILWSQPPEDTQNGMVQYKVCWCLGNTRYHINKT--VDGST	834
Db	665	SAAPQNLSEVR--NSKSIYIFWQPPSSTQNGQITGYKIRYKASKRSDVTTETLVGTQ	722
Qy	835	FSVVIPLVPGIRYSVEVAASTGAGSGVKSE-----PQFIQLDAHGNPV	878
Db	723	LSQLEIGLDRGTEYFNRAALTIVNGTGPAIDWLSAETFESDLDETRVPE--VPSSLHVRPL	781
Qy	879	-----SPEDQVSLAQQISDVVKQPAFIAGIACAACWILMVFSIWLYRHRKKRGLT	929
Db	782	VTSIVVSWTPPENQ-----NIVVRGYAIGYGISGPHAQTIKVD-----YKQR-----	823
Qy	930	STYAGIRKV--PSFTTPTV--TYQRGGEAV---SSGGVP-----GLLNISEP--AAQPW	975
Db	824	--YTTIENLDPSHHYVITLKAFFNNVGEGIPLYESAIVRPHDTSEVDLFVINAPYTPVPD	881
Qy	976	LADTWNTG----NNHNDCSI SCCTAGNGSDSNLTYSRPAACIANYNQLDNKQTNL	1030
Db	882	PTPMPMPVGVAQISLSDHTIRITW-----ADNSLPKHQKITD--SRITYV--RWKTN--	929
Qy	1031	MLPESTVYGDVDSLKNINEMKTFNSPN-----LKDRFVNPSSGQPTPYATTQLIQSN	1082
Db	930	--IPANTKYKNAN--ATTLSYLVTLGKPNLTLYEFSVMVTKGRSSSTWSMTAHGATFELVPTS	987
Qy	1083	LSNNMNNNGSGDSGEK---HWKPLGQQKQEV-----APVQYNIVEQ---NKLN	1123
Db	988	PPKDVTVVSKGKPRTIIVNWQPPSEANGKITGYIIYSTDVNAEIHWDWIEPVVGNRLT	1047
Qy	1124	KDYRANDTVPPTIIPYNSQYDQNTGG-----SYNSSDR--GSSTSGSQGHKKG	1168
Db	1048	--HQIQLTLDTPYFYFKIQARNSKGMGPSEAVQFRTPKADSSDKMPNDQALGSAG--KG	1103
Qy	1169	ARTPKVPKQGGMNWADLPPPPAHPPPHNSSE---EYNI-----	1204
Db	1104	SRLPDL-----GSDYKPMMSGNSPHGSPTSPLDSNMLLVIIVSVGITIVVVVVI	1156
Qy	1205	-----SVDESVDQEMPC-PVPPARMYLQQDELEEEEDERGPTP--PVR	1244
Db	1157	FCTRRTTSQKKKRAACKSVNGSHYKGNCKDKVPPDLWIHHERLELKPIDKSPDPNPV-	1215
Qy	1245	GAASSPAAVSYSHQSTAILTPSPQBELQPMQDCPEETHMQ-----HQPD-----	1290
Db	1216	-----MTDPTIPRNSQDITP--VDNSMDSNIHQRRNSYRGHESEDMSMTLAG	1260
Qy	1291	-RRRQVSPPPP---PPRPISPPHTYIGISGPLVSDMDTDAPEEEEDADMEVAKMQTRR	1345
Db	1261	RRGMRPKMMMPFDSQPPQPVISAH-----PIHS--LDNPHHHFHSSSL-----	1301
Qy	1346	LLLRGLEQTPA-----SSVGDLESSVTGSMINGWGSASEEDNISSGRSSVSSSDGSFF	1398
Db	1302	-----ASPARSHLYHPSSPWPIGTSM--SLSDRANSTESVRNTPSTDTPASSSQTCC	1352
Qy	1399	TDADFAQAVAAAAEYAGLKVARROMQDAAGRRHFHASQCPRPTSPVSTDSNMSAAVMQKT	1458


```

ID Q90610      PRELIMINARY;      PRT; 1443 AA.
AC Q90610;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE NEOGENIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI TaxID=9031

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:19:37 ; Search time 233.01 Seconds
(without alignments)
63.689 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 QIVAQGRITVTFPCETKGNPQ.....TSAALSQSQRPRPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2276	99.8	434	20	Y13567	Human Robo 2 polyp
2	2276	99.8	434	20	Y08405	Human partial ROBO
3	914	40.1	1649	20	Y08404	Human ROBO1 protei
4	913	40.0	1651	20	Y13566	Human Robo 1 polyp
5	877.5	38.5	753	20	W83927	Human T85 protein.
6	561.5	24.6	1297	20	Y13565	C. elegans Robo po
7	561.5	24.6	1297	20	Y08403	C. elegans ROBO pr
8	545	23.9	1395	20	Y13563	Drosophila Robo 1
9	545	23.9	1395	20	Y08401	Drosophila sp. ROB
10	473.5	20.8	1380	20	Y08402	Drosophila sp. ROB
11	473.5	20.8	1381	20	Y13564	Drosophila Robo 2
12	324.5	14.2	1911	16	R71726	Human PTP-OB. Hom

13	324.5	14.2	1911	18	W27225	Human protein tyro
14	324.5	14.2	1911	20	W94027	Human protein tyro
15	323.5	14.2	1501	16	R72858	Rat receptor type-
16	308.5	13.5	1897	21	Y81785	Human protein tyro
17	308.5	13.5	1897	21	Y56100	LAR tyrosine phosph
18	305.5	13.4	761	17	R92255	Neural cell adhesi
19	285.5	12.5	1192	19	W57900	Protein of clone C
20	285.5	12.5	1299	21	Y40439	Human Nr-CAM prote
21	285.5	12.5	1496	20	W81030	Melanoma associate
22	285.5	12.5	1496	21	Y70469	Human p53 target m
23	276.5	12.1	582	17	R92256	Neural cell adhesi
24	268	11.8	1257	20	W74152	Human L1 cell adhe
25	267.5	11.7	1028	19	W29667	Homo sapiens DL185
26	264.5	11.6	1242	19	W52287	Rattus norvegicus
27	262.5	11.5	1304	19	W59994	Human neural cell
28	254	11.1	868	17	R92717	Mouse muscle-local
29	253	11.1	869	18	W26611	Human muscle-speci
30	253	11.1	869	18	W26506	Human Dmk receptor
31	250.5	11.0	867	19	W62583	Mouse receptor tyr
32	250.5	11.0	871	17	R84087	Nsk2 receptor. Mu
33	250.5	11.0	871	19	W62568	Mouse receptor tyr
34	250.5	11.0	881	17	R84091	Nsk2 receptor with
35	250.5	11.0	881	19	W62572	Mouse Nsk2 (altern
36	250	11.0	1091	18	W41641	Sequence used in d
37	250	11.0	1091	20	Y08010	Mouse LIG-1 protei
38	249	10.9	860	17	R92716	Mouse muscle-local
39	249	10.9	1447	16	R68553	Deleted in colorec
40	249	10.9	1447	20	Y33498	Human DCC protein.
41	249	10.9	1728	12	R13144	Deleted in Colorec
42	247.5	10.9	1018	15	R63759	Human contactin (E
43	247.5	10.9	1018	17	R87028	Human contactin.
44	245.5	10.8	863	17	R84088	Nsk2 receptor with
45	244.5	10.7	863	19	W62569	Alternatively spli

ALIGNMENTS

RESULT 1
Y13567
ID Y13567 standard; Protein; 434 AA.
XX
AC Y13567;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 285
FT /label= unknown
FT /note= "encoded by GTN"
FT Misc-difference 396
FT /label= unknown
FT /note= "encoded by NTT"
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX

DR WPI; 1999-338008/28.
DR N-PSDB; X55771.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 49-50; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commisured) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 434 AA;

Query Match 99.8%; Score 2276; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.1e-144;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIVAGQRTVTFPCETKGNPQPAVFWQEGSQNLLFPNQPPQNSRCSVSPGTDLTITNIQ 60
Db 1 qivagqrtvtfcetkgnppavfwqegsqnllfpnqpqnsrscsvsptgdlititniq 60
Qy 61 RSDAGYYICQALTAVAGSILAKAQLEVTDLVTRPPPIILQGPANQTLAVDGTALLKCKAT 120
Db 61 rsdagyyicqaltvagsilakaglevtldvtrpppiilqgpangtlavdgtallkckat 120
Qy 121 GDPLPVISWLKEGFTFPPGRDPRATIEQGLTIQIKNLRIISDTGTYTCVATSSSGEASWSAV 180
Db 121 gdplpviswlkegftfppgrdpratieqegtlqiknlrisdtgtytcvatsssggeaswsav 180
Qy 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLQSWQPGTPTGLPASAYIIIEAFSQ 240
Db 181 ldvtesgatisknydlsldpgppskpqvtdvtnsvtlswqpgtptglpasayileafsq 240
Qy 241 SVNSWQTVANHVKTITLYTVRGLRPNTIYLFMVRINPKVSVTQXPKQKNGSTWANVPL 300
Db 241 svnswqtvanhvkttlytvrgrlpntiylfmvrainpkvsvtqxpkqkngstwanvpl 300
Qy 301 PPPVQPLPGTELEHYAVEQENGYSWCPPLPVQTYLHQGLEDELEEDDDRVPTPPV 360
Db 301 pppvqplpgtelehyaveqengydsdswcpplpvqtylhqgledeleedddrvptppv 360
Qy 361 RGVASSPAISFGQQSTATLTSPREEMQPMQASXPFTSSQRPRTPSPFSTDSNTSAALS 420
Db 361 rgvasspalsfgqqstatltspreemqpmqaspxftssqrprtpspfstdsntsaaals 420
Qy 421 QSQRPRPTKKHKG 434
Db 421 qsqrprptkhhkgg 434

RESULT 2

Y08405

ID Y08405 standard; Protein; 434 AA.

XX

AC Y08405;

XX

DT 24-JUL-1999 (first entry)

XX

DE Human partial ROBO2 protein.

XX

KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.

XX

OS Homo sapiens.

XX

PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57254.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 72-73; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 434 AA;

Query Match 99.8%; Score 2276; DB 20; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.1e-144;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIVAGQRTVTFPCETKGNPQPAVFWQEGSQNLLFPNQPPQNSRCSVSPGTDLTITNIQ 60
Db 1 qivagqrtvtfcetkgnppavfwqegsqnllfpnqpqnsrscsvsptgdlititniq 60
Qy 61 RSDAGYYICQALTAVAGSILAKAQLEVTDLVTRPPPIILQGPANQTLAVDGTALLKCKAT 120
Db 61 rsdagyyicqaltvagsilakaglevtldvtrpppiilqgpangtlavdgtallkckat 120
Qy 121 GDPLPVISWLKEGFTFPPGRDPRATIEQGLTIQIKNLRIISDTGTYTCVATSSSGEASWSAV 180
Db 121 gdplpviswlkegftfppgrdpratieqegtlqiknlrisdtgtytcvatsssggeaswsav 180
Qy 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLQSWQPGTPTGLPASAYIIIEAFSQ 240
Db 181 ldvtesgatisknydlsldpgppskpqvtdvtnsvtlswqpgtptglpasayileafsq 240
Qy 241 SVNSWQTVANHVKTITLYTVRGLRPNTIYLFMVRINPKVSVTQXPKQKNGSTWANVPL 300
Db 241 svnswqtvanhvkttlytvrgrlpntiylfmvrainpkvsvtqxpkqkngstwanvpl 300
Qy 301 PPPVQPLPGTELEHYAVEQENGYSWCPPLPVQTYLHQGLEDELEEDDDRVPTPPV 360
Db 301 pppvqplpgtelehyaveqengydsdswcpplpvqtylhqgledeleedddrvptppv 360
Qy 361 RGVASSPAISFGQQSTATLTSPREEMQPMQASXPFTSSQRPRTPSPFSTDSNTSAALS 420
Db 361 rgvasspalsfgqqstatltspreemqpmqaspxftssqrprtpspfstdsntsaaals 420
Qy 421 QSQRPRPTKKHKG 434
Db 421 qsqrprptkhhkgg 434

RESULT 3

Y08404

ID Y08404 standard; Protein; 1649 AA.

XX

AC Y08404;

```

RESULT      4
Y13566
ID   Y13566 standard; Protein; 1651 AA.
XX
AC   Y13566;
XX
DT   30-JUL-1999   (first entry)
XX
DE   Human Robo 1 polypeptide.
XX
KW   Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW   modulation; nerve cell function.
XX
OS   Homo sapiens.
XX
PN   WO9925833-A1.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-US24327.
XX
PR   14-NOV-1997;   97US-0065543.
XX
PA   (REGC ) UNIV CALIFORNIA.

```

Qy 298 297

Db 1436 qcprptspvstdsnmsaavmqktrpakklkhpg 1469

FT /note= "cytokine receptor homology N-terminal

XX
PN WO9848051-A2.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-US07714.
XX
PR 10-OCT-1997; 97US-0062017.
PR 18-APR-1997; 97US-0044746.
XX
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI Holtzman D, McCarthy SA;
XX
DR WPI; 1999-024021/02.
DR N-PSDB; V69278.
XX
PT New isolated human FTHMA-070 and T85 proteins - used to develop
PT products for the diagnosis and therapy of disorders involving
PT cellular processes, e.g. neuronal development.
XX
PS Claim 31; Fig 3; 127pp; English.
XX
CC This is the amino acid sequence of a novel human protein designated
CC T85, and also referred to as FMHB-6D4 and FMHB-SD4. T85 cDNA (see
CC V69278) was identified in a human foetal brain cDNA library using a
CC screen designed to identify genes encoding proteins having a
CC functional signal sequence. T85 nucleic acids and polypeptides of
CC the invention are useful as modulating agents in regulating a
CC variety of cellular processes. They can be used for identifying
CC compounds which bind to or modulate the activity of the polypeptides
CC (claimed). They can also be used in screening assays, detection
CC assays (e.g. chromosomal mapping, tissue typing, forensic biology),
CC predictive medicine (e.g. diagnostic assays, prognostic assays,
CC monitoring clinical trials, and pharmacogenomics), and methods of
CC treatment (e.g. therapeutic and prophylactic) e.g. for neurological
CC disorders.
XX
SQ Sequence 753 AA;

Query Match 38.5%; Score 877.5; DB 20; Length 753;
Best Local Similarity 51.1%; Pred. No. 9.4e-51;
Matches 180; Conservative 48; Mismatches 93; Indels 31; Gaps 6;

Qy 1 QIVAQGRVTVPFCETKGNPQPAVFWQKEGSQLNLFNPQPPQNSRCSVSPTGDLTITNIQ 60
Db 324 qvalgrvtvfqceatgnppaifwrrgsqnlfsyqppqssrsfvsqtdlittinvq 383

Qy 61 RSDAGYYICQALTAVGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
Db 384 rsdvgyycqtlnvagsiitkaylevtdviadrpppvirgppvngtvdgtfvlscvat 443

Qy 121 GDPLPVISWLKEGFTFPGRDPRATIQEQTLQIKNLRISDTGTYYTCVATSSSGEASWSAV 180
Db 444 gspvptilwrkdglvstqdsrikqlengylqiryaklgdtgrytciaastpsgeatwsay 503

Qy 181 LDVTESGATIS--KNYDLDLPGLPSPKQVTDVTKNSVTLWSQPGTPTGLPASAYIIIEAF 238
Db 504 ievgefgyvpprptdpnlpsapskevtdvsrntvltswqpnlnsgatptsyieaf 563

Qy 239 SQSVNSWQTVANHVKTTLTYVRGLRPNTIYLFMVRAIN----- 277
Db 564 shasgsswgtvaenvktetsaiklknaiylfivraanaygisdpqisdvpktqdvlp 623

Qy 278 PKVSVTQXKPKQKNGSTWANVPLPPVPQLPGTELE-HYAVEQQE---NGY 325
Db 624 tsqgvdhkqvqrelgn--avhlhnpvtv--lssssievhtwcvdqgsqyiggy 671

RESULT 6
Y13565
ID Y13565 standard; Protein; 1297 AA.

XX
AC Y13565;
XX
DT 30-JUL-1999 (first entry)
XX
DE C. elegans Robo polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Caenorhabditis elegans.
XX
PN WO9925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55769.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 38-39; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1297 AA;

Query Match 24.6%; Score 561.5; DB 20; Length 1297;
Best Local Similarity 37.2%; Pred. No. 1.9e-29;
Matches 124; Conservative 48; Mismatches 114; Indels 47; Gaps 6;

Qy 1 QIVAQGRVTVPFCETKGNPQPAVFWQKEGSQLNLFNPQPPQNSRCSVSPTGDLTITNIQ 60
Db 325 qsvpaggtatfectlvqpspayfwskeggqdlfipsy-vsadgrtkvsptgtltievr 383

Qy 61 RSDAGYYICQALTAVGSILAKAQLV-----V 86
Db 384 qvdegayvcagmnsagsslskaalkatfetkgrvqkkkskmgkqkqknvgsiikylisav 443

Qy 87 TDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDPLPVISWLKEGFTFPGRDPRATIQ 146
Db 444 tgntpakppptiehqngqtlmvgssailpcqasgkptpgiswlrldgidtdsrisqh 503

Qy 147 EQTLQIKNLRISDTGTYYTCVATSSSGEASWSAVLDVTE--SGATISKNYDLDLPGLPSPK 204
Db 504 stgslhiadlkkpdtgvytciaknedgestwsasltvedtwsnaqfvmpdpnsfsspt 563

Qy 205 KPQVTDVTKNSVTLWSQ-PGTPTGLPASAYIIIEAFSQSVNSWQTVANHVKTTLTYVRGL 263
Db 564 qpilvnvtdtevelhwnapstsgagpityliqyyspdlggtwnipdyvasteyrikgl 623

Qy 264 RPNTIYLFMVRAIN-----PKVS---VTQXKP 287
Db 624 kpshsymfviraenekgigtspvssalvttskp 656

Query Match 24.6%; Score 561.5; DB 20; Length 1297;
Best Local Similarity 37.2%; Pred. No. 1.9e-29;
Matches 124; Conservative 48; Mismatches 114; Indels 47; Gaps 6;

Query Match 23.9%; Score 545; DB 20; Length 1395;
Best Local Similarity 40.9%; Pred. No. 2.6e-28;
Matches 115; Conservative 48; Mismatches 104; Indels 14; Gaps 7;

```
Qy      9 VTFPCETGPNQPAVFWQEGSKSNLLFPNQPPQNSCRSVSPGTDLTITNIQRSDAGYII 68  
       | || ||| |:||| ||| |:||| : | |: | ||::: | |||:  
Db     362 vqlpcmasgnpppsvfwtkegvstlmfpn---sshgqyyvaadgtlqitdrvqedegyvv 418  
  
Qy     69 CQALTAVAGSILKAQLEVTDLVLDRPPIIILQGPNQTLAVDGTALLKCKATGDPLPVIS 128  
       | |: | |: |: |: |: ||||| ||||| | |: |||: | | |  
Db    419 csafsvdvsstvyrlqvwssv-derpppiilqipganqlpkgsvatlpccratgnpspri 477  
  
Qy    129 WLKEGFTFPFGEDPRATIOEQGLTIQKNLRISDGTYYTCVATSSSGEASWSAVLVDTESGA 188  
       | |: | |: |: |: |: ::||: ||||| |: | |||: | | |: |  
Db   478 wfhdghavgaqn-rysiigsgslrvddllqldsgtytctasgergetsaaatlvekpqs 536  
  
Qy    189 T-ISKNYDLSDLPGPPSPKQVTDVTKNSVTLW----QPGETGPLPASAIIEAFSQSV 242  
       | |: | |: |: |: |: ::||: ||||| |: | |||: | | |: |  
Db   537 tslhraadspstyappgtpkvlnvrtsislrwaksqekpgavg--pilgytvveyfspd 594  
  
Qy    243 SNSWQTIVANHVKTTLTYVRGLRPNTIYLEMVRAINPK-VSV 282
```

```

Qy      9 VTFPCETKGNPQAVFWQKEGSQNLFFPNQPPQPNCSRCSVSPGTDLTITNIQRSDAGYII 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    362 vqlpcmasggnpppsvfwtkegvstlmfpn---sshgrqyyvaadgtlqitdvrqdeggyvv 418

Qy     69 CQALTVAGSILAKAQLEVTDLVTRDPPPIILIQGPANQTLAVDGTALLKCKATDGPLVIS 128
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    419 csafsvdsstsvrflqyvsav-derpppiilqigpanqlpkgsvatlpcratgnpsprk 477

Qy    129 WLKEGTFPPGRDPRATIEQGTLOIKNLRISDTGYTTCVATSSSGEASWSAVLDVTESGA 188
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    478 wfhdghavqagn-rysiigssrlrvddlqlsdsgtytctasgergetswaatltvekpgs 536

Qy    189 T-ISKNYDLSDLPGPPSKPQVTDVTKNSVTLGW----QPGETPTLPASAYII EAFSQSV 242
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    537 tshhraadstypappgtpkvlinvrtsislrwaksqekpgavg--piilgytveyfspdl 594

Qy    243 SNSQVTQVANHVKTTTLTYVRGLRPNTIYLEFMVRAINPK-VSV 282
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    595 gtgwivaahrvadtavtisltpatsvflvrtaentqaisv 635

```

RESULT 10
Y08402
ID Y08402 standard; Protein; 1380 AA.
XX
AC Y08402;
XX
DT 24-JUL-1999 (first entry)
XX
DE Drosophila sp. ROBO2 extracellular domain protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Drosophila sp.
XX
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22154.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57251.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 52-56; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1380 AA;

Query Match 20.8%; Score 473.5; DB 20; Length 1380;
Best Local Similarity 28.6%; Pred. No. 1.5e-23;
Matches 134; Conservative 66; Mismatches 182; Indels 87; Gaps 15;

```

Qy      1 QIVAQGRVTVPFCCTKGNPQAVFVQKEGSONLLFPNQPPQPNSCRVSPTGD--LTITN 175
      |:| | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db    306 qlveigdevlfecqanghprptlywsvegnslllpgy-rdgrmevtltpgrsvlsiar 364

Qy      59 IQRSAGYIYI-CQALTVAGSILAKAQLEVTDLVLTDRPPPIILQGPANQTLAVDGTALLKC 117
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db    365 faredsgkvvtcnaInavgsvssrtvsv-dtqfelpppfielqgpgvngtlpvksivlpc 423

Qy     118 KATGDPLPVISWLKEGTFPPGRD-PRATIQEQGTQLQIKNL-RISDTGTYTCVATSSSGEA 175
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     424 rtlgtppqgswyldgpidvqehernrlsdagaltisdrlqrhedeglytcvasnrngks 483

Qy     176 SWSAVLDV---TESGATISKNYDLSDLGPSPKQVTDVTKNSVTLWS-QPGTPGTLPAS 231
      ||| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     484 swsgylrldttnpnikffrapelstypgppgkpmvekgensvtlswtrsnkvvgsslv 543

Qy     232 AYIIIEAFSQSVNSQWTVANHVKTTLYTVRGLRPNTIYLEFMVRAINPKVSVTXKPKQKN 291
      |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db     544 gyviemfgknetgdwvavgtvrqntfttqgllpgvnyffliraensh----- 591

Qy     292 GSTWANVPLPPPPVQLP-GTELEHYAVEQQE-----NGYSDSDSWCPLPV 336

```

```
Qy      1 QIVAQGRVTTFPCETKGNQPAVFWQKEGSQNLLFPNQPPQPNRSRCSVSPGTD--LTIITN 58
      ||| ||| ||| :|:|: ||| ||| : ||| ||| : ||| |||
Db     306 qlweigdevlfecqanghprptlywsvegnsillpgy-rdgrmevltlpegrsvlsiar 364

Qy     59 IQRS DAGYI -CQALTVAGSILAKAQLEVTDLVDRPPPIILQGPAQT LAVDGTALLKC 117
      ||| : ||| ||| : : ||| : ||| ||| ||| ||| : |||
Db     365 faredsgkvvtcnaalnavgsvssrtvsvs-dtqfelpppiileggpvnqtlpvksilvlp 423

Qy    118 KATGDPLPVISWLKEGFTTFPGRD-PRATIQEGTLQIKNL-RISDTGYTTCVATSSSGEA 175
      ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 12
R71726
ID R71726 standard; Protein; 1911 AA.
XX
AC R71726;
XX
DT 17-OCT-1995 (first entry)
XX
DE Human PTP-OB.
XX
KW PTP-OB; protein tyrosine phosphatase; osteoblast; differentiation;
KW osteoclast; osteoporosis; bone; cancer; osteosarcoma.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..29
FT /label= Sig_peptide
FT Modified-site 250
FT /label= N-glycosylation_site
FT Modified-site 721
FT /label= N-glycosylation_site
FT Modified-site 919
FT /label= N-glycosylation_site
FT Domain 1253..1277
FT /label= Extracellular_domain
XX
PN WO9507935-A.
XX
PD 23-MAR-1995.
XX
PF 09-SEP-1994; '94WO-US10166.
XX
PR 14-SEP-1993; '93US-0122032.
XX
PA (MERI ) MERCK & CO INC.
XX
PI Rodan GA, Rutledge SJ, Schmidt A;
XX
DR WPI; 1995-131318/17.
DR N-PSDB; Q86478.
XX
PT Protein tyrosine phosphate protein PTP-OB specifically expressed
PT in bone cells - modulators of which are used to treat, e.g.
PT osteoporosis, and prevent and treat bone loss and cancer.
XX
PS Claim 1; Page 44-45; 63pp; English.
XX
CC PCR amplification of cDNA derived from human osteosarcoma
CC Saos-2/B10 using primers based on conserved regions of protein

```

[illegible]

[illegible]

```

Db      213 agvrysspanlyvrvvrapfsilpmshelpggnvnitcavagsspmyvkwmqgaedl 272
Qy      180 -----VLDVTESGATISKNY-----DLSDLPGPPSPKQVTD 210
          ||::| | | |
Db      273 tpeddmpgvgrnvleldt--vkdsanyhpcvamsslglvieavaqitkslpkagptpmvte 330
Qy      211 VTKNSVTLSWQPTGTLPASAYIEAFSSQSNSWQTVANHVKTTLTYVRGLRPNTIYL 270
          |:|:| | :| | | | | | | | | | | | | | | | |
Db      331 ntatsititwdsngpd--pvsyyvyiekysqdgpqy-ikedittrsysigslpnseye 387
Qy      271 FMVRAINPKVSVTQXKPQNNGS-TWANVPLPPP-----PVQP--- 307
          | | | | : | : | | | | | | | | | |
Db      388 iwsavn---sigggppsesvvtregaparprnvqarmisattmivgweepvepngl 444
Qy      308 LPGTELEHYAVEQQENGYSDSWSCPLLVPQTYLHQGLEDD-----ELEEDD----- 352
          : | : | : | : | : | : | : | : | : | : | : |
Db      445 irgyrv-yytme-----pehpvgwnqkhnvddsltttgvsllledetytvrvla 491
Qy      353 -----DRVPTPPVRGVASSP---AISFGQQSTATLTSPSR----- 384
          | : : | | | : | : | |
Db      492 ftsvgdgplsdpiqvktqggvgpggmnlraearsetsitlswsprrgesiiykellfreg 551
Qy      385 -----EEMQP-----MLQASP---XFTSSQRPR-----PTSP-- 408
          |::| | : | | | | | | | | | |
Db      552 dhgrevgtrfdpttsyvvedlkpnteyafriaarspgglgaftpvvrqrqlskspappq 611
Qy      409 -FSTDSNTSAALSQSQRPRPTKKHKG 433
          | | | | | | | | | |
Db      612 dvkcvsrvrstaillswrppppethng 637

```

RESULT 15
R72858
ID R72858 standard; Protein; 1501 AA.
XX
AC R72858;
XX
DT 21-NOV-1995 (first entry)
XX
DE Rat receptor type-protein tyrosine phosphatase sigma.
XX
KW Receptor type tyrosine phosphatase sigma; cell; differentiation;
KW metabolism; cell cycle; behaviour; motility; contact inhibition;
KW virus; inflammation; cellular transformation; cancer;
KW neuroblastomas; antibody; detection; quantification.
XX
OS Rattus rattus.
XX
PN W09509656-A.
XX
PD 13-APR-1995.
XX
PF 30-SEP-1994; 94WO-US11163.
XX
PR 01-OCT-1993; 93US-0130570.
XX
PA (UUNY) UNIV NEW YORK STATE.
XX
PI Schlessinger J, Yan H;
XX
DR WPI; 1995-155068/20.
DR N-PSDB; Q86902.
XX
PT Novel, isolated receptor-type protein tyrosine phosphatase-sigma
PT - and encoding DNA, useful e.g. for detecting neuro-blastomas
XX
PS Claim 2; Figure 2; 105pp; English.
XX
CC Ligands binding to the receptor-type protein tyrosine phosphatase
CC sigma (RPTP sigma) protein may be used as drugs to modulate cellular
CC processes, such as differentiation, metabolism and cell cycle
CC control, and cellular behaviour such as motility and contact
CC inhibitions. In addition they may affect abnormal or potentially

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:16 ; Search time 325.28 Seconds
(without alignments)
901595 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 QIVAQGRVTVPCECTKGNPQ.....TSAALSQSQRPRPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	911	40.0	1612 2	T30805 duttl protein - mo
2	911	40.0	1651 2	T14160 transmembrane rece
3	755.5	33.1	1344 2	T14316 rig-1 protein - mo
4	579	25.4	1273 2	T42405 sax-3 protein - Ca
5	388	17.0	874 2	T29548 hypothetical prote
6	331.5	14.5	2222 12	T13924 sdk protein - frui
7	326	14.3	1907 2	S50893 protein-tyrosine-p
8	325	14.3	1499 2	I50212 protein-tyrosine-p
9	323.5	14.2	1501 2	I58148 protein-tyrosine-p
10	323.5	14.2	1863 2	S46217 protein-tyrosine-p
11	317	13.9	1894 2	C54689 protein-tyrosine-p
12	308.5	13.5	1897 1	TDHULK leukocyte antigen-
13	307.5	13.5	1912 2	A56178 protein-tyrosine-p
14	305.5	13.4	1898 2	S46216 leukocyte antigen-
15	297.5	13.0	1232 2	T43027 neural cell adhesi
16	296.5	13.0	1277 2	T30532 neural cell adhesi
17	295.5	13.0	2029 1	TDFFLK protein-tyrosine-p
18	292.5	12.8	1262 1	B48758 protein-tyrosine-p
19	292.5	12.8	1496 1	A48758 protein-tyrosine-p
20	282	12.4	1443 2	I50600 neogenin - chicken
21	276	12.1	1028 2	I58164 BIG-1 protein - ra
22	274	12.0	1197 2	T30581 neural cell adhesi
23	269	11.8	1028 2	A53449 plasmacytoma-assoc
24	268	11.8	1257 1	A41060 neural cell adhesi
25	266.5	11.7	1427 2	I51669 tumor suppressor -
26	265.5	11.6	1260 1	S05479 neural cell adhesi
27	264.5	11.6	1256 2	T03096 CDO protein - rat
28	262.5	11.5	1272 2	S26180 neurofascin - chic
29	261.5	11.5	1259 2	S36126 neural cell adhesi

30	258	11.3	1239	1	A32579	neuroglian - fruit
31	256.5	11.2	423	2	T29549	hypothetical prote
32	256	11.2	1040	2	A49356	transient axonal g
33	255.5	11.2	1375	2	T13822	frazzled gene prot
34	253	11.1	1018	2	JC4211	neural adhesion pr
35	252.5	11.1	1259	2	A43425	Bravo/Nr-CAM cell
36	251.5	11.0	1036	2	S22383	axonin 1 precursor
37	251	11.0	761	1	IJHUNG	neural cell adhesi
38	251	11.0	1880	2	T18531	tractin - medicina
39	250.5	11.0	871	1	I48696	protein-tyrosine k
40	250.5	11.0	881	1	I48697	protein-tyrosine k
41	250.5	11.0	1268	1	A39640	neural cell adhesi
42	250	11.0	1091	2	A58532	glial cell membran
43	249	10.9	1447	2	A54100	tumor suppressor p
44	248.5	10.9	1437	2	T31093	probable protein-t
45	248	10.9	725	1	IJMSNG	neural cell adhesi

ALIGNMENTS

RESULT 1

T30805

dutt1 protein - mouse

N;Alternate names: transmembrane receptor protein Robol homolog

C;Species: Mus musculus (house mouse)

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C;Accession: T30805

R;Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.

submitted to the EMBL Data Library, July 1998

A;Description: The mouse homologue of human DUTT1/H-robol gene: protein sequence and

A;Reference number: Z20879

A;Accession: T30805

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-1612 <NUM>

A;Cross-references: EMBL:Y17793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1

A;Experimental source: brain

C;Genetics:

A;Gene: duttl1

A;Map position: 16

Query Match 40.0%; Score 911; DB 2; Length 1612;
Best Local Similarity 22.8%; Pred. No. 1.7e-46;
Matches 254; Conservative 55; Mismatches 120; Indels 686; Gaps 10;

Qy 1 QIVAQGRVTVPCECTKGNPQPAVFWQKEGSQLNLFPPNQPPNSRCSVPTGDLTITNIQ 60

Db 321 QVVALGRVTVPCEATGNPQPAIFWRREGSQLNLFVYQPPQSSRSVSPTGDLTITNVQ 380

Qy 61 RSDAGYYICQALTAVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120

Db 381 RSDVGGYYICQTLNVAGSIITKAYLEVTVDVIADRPPIVIRQGPVNQTVAVDGTLLSCVAT 440

Qy 121 GDPLPVISWLKEGFTFPGDRPRATIEQGTILQIKNLISDTGTYTCVATSSSGEASWSAV 180

Db 441 GSPAPTILWRKDGVLVSTQDSRIKQLESGVLQIRYAKLGDGTCTASTPSGEATWSAY 500

Qy 181 LDVTESGATIS--KNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTPTGLPASAYIIIEAF 236

Db 501 IEVQEGVPVQPPRPTDNLIPSAPSKPEVTDVSKNTVTLSWQPNLNGATPTSYIIIEAF 560

Qy 239 SOSVSNSWQTVANHVKTTLYTVRGLRPNTIYLFMVRAIN----- 277

Db 561 SHAGSSSQTAENVKIETFAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTDVPP 620

Qy 278 ----- 277

Db 621 TSQGVDHQVQRELGNVVLHLNPTILSSSSVEVHWTVQQSQVIQYKILYRPSGASHG 680

Qy 278 -----PK 279

Query Match 40.0%; Score 911; DB 2; Length 1651;
Best Local Similarity 23.0%; Pred. No. 1.8e-46;

[illegible]

Qy 401 QRPRTSPFSTDSNTSALSQSQRPRPTKKHKG 434
| | | | | | | | | | | | | | | | | | | | | |
Db 1436 QCPRTSPVSTDSNMSAAVIQKARPTKKQHPG 1469

RESULT 3
T14316
rig-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14316
R:Yuan, S.S.F.; Cox, L.A.; Dasika, G.K.; Lee, E.Y.H.P.
submitted to the EMBL Data Library, April 1998
A:Reference number: Z17975
A:Accession: T14316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1344 <YUA>
A:Cross-references: EMBL:AF060570; NID:g4206385; PID:g4206386; PIDN:AAD11628.1

Query Match 33.1%; Score 755.5; DB 2; Length 1344;
Best Local Similarity 39.6%; Pred. No. 2.4e-37;
Matches 176; Conservative 55; Mismatches 141; Indels 73; Gaps 9;

Qy 1 QIVAGRTVTFPCETKGNQPAVFWQEGSQNLLFPNQPPNSRCSVSPTGDLTITNIQ 60
| | | | | | | | | | | | | | | | | | | | | |
Db 334 QTVAGANVSFQCEKGNPPAIFWQEGSQVLLFPSSQLPMGRLLVSPRGQNLNITEVK 393

Qy 61 RSDAGYICQALTIVAGSILAKAQLVTDVLRPPPIIILQGPANQTLAVDGTALLKCKAT 120
| | | | | | | | | | | | | | | | | | | | | |
Db 394 IGDGGYVYQAVSVAGSILAKALLEIKGASIDGLPPIIILQGPANQTLVLGSSVWLPORVI 453

Qy 121 GDPLPVISWLKEGFTFPGDRPRATIEQGTILQIKNLRIISDTGTYTCVATSSSGEASWSAV 180
| | | | | | | | | | | | | | | | | | | | | |
Db 454 GNPQNIQWKKDERWLGDDSQNLMDNGTLHIAISQEMDGFYSCVAKSSIGEATWNSW 513

Qy 181 LDVTES-GATISKNYDLSDLPGPSKPQVTDVTKNSVTLSWPGTPTGLPASAYIEAFS 239
| | | | | | | | | | | | | | | | | | | | | |
Db 514 LRKQEDWASPGPATGPNPPGPSQPIVTEVNTSITITWKNPQSGATATSYVIEAFS 573

Qy 240 QSVNSNSWQTVANHVKTLYTVRGLRPNTIYLFMVRIN-----PKVSVTQXKPKNNGSTW 295
| | | | | | | | | | | | | | | | | | | | | |
Db 574 QAAGNTWRTVADGVQLETITISGLQNTIYFLVRVAGAWLSESPVSEPVQTDSS-- 631

Qy 296 ANVPLPPPPVQLPGTE-LEHYAVEQQE-----N 323
| | | | | | | | | | | | | | | | | | | | | |
Db 632 ----LSRPAEDPWKQRLAEVAVRMQEPVTLGPRTLVQSVTVQVQLVQGFVRSWRIA 687

Qy 324 GYDSDSW-----CPP-----LPVQTYLHQGLEDE-----LEEDDRVPT 357
| | | | | | | | | | | | | | | | | | | | | |
Db 688 GLDQGSWMLDLQSPHKQSTVLRGLPPGAQIQIKVQVQEGGLGAESPFTVRSIPEEAPS 747

Qy 358 PPVRGVASSPAISFGQSTATLTPS 382
| | | | | | | | | | | | | | | | | | | | | |
Db 748 GPPQGV----AVALGGDRNSSVTYS 768

RESULT 4
T42405
sax-3 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42405
R:Zallen, J.A.; Yi, B.A.; Bargmann, C.I.
Cell 92, 217-227, 1998
A:Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple asp
A:Reference number: Z22160; MUID:98117250
A:Accession: T42405
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <ZAL>

A:Cross-references: EMBL:AF041053; NID:g2804779; PIDN:AAC38848.1; PID:g2804780
C:Genetics:
A:Note: sax-3
C:Function:
A:Description: sax-3 function is required at the time of axon guidance

Query Match 25.4%; Score 579; DB 2; Length 1273;
Best Local Similarity 40.8%; Pred. No. 6.8e-27;
Matches 124; Conservative 48; Mismatches 114; Indels 18; Gaps 6;

Qy 1 QIVAGRTVTFPCETKGNQPAVFWQEGSQNLLFPNQPPNSRCSVSPTGDLTITNIQ 60
| | | | | | | | | | | | | | | | | | | | | |
Db 326 QSVFAGGTATFECTLVGQSPAYFWSKEGQDILLFSPY-VSADGRTKVSPTGLTIEEVR 384

Qy 61 RSDAGYICQALTIVAGSILAKAQLVTDVLRPPPIIILQGPANQTLAVDGTALL 115
| | | | | | | | | | | | | | | | | | | | | |
Db 385 QVDEGAYVCAGMNSAGSSLSKALKVTTKAVTGNTPAKPPPTIEHGHQNTLMVGSAIL 444

Qy 116 KCKATGDPLPVISWLKEGFTFPGDRPRATIEQGTILQIKNLRIISDTGTYTCVATSSSGEA 175
| | | | | | | | | | | | | | | | | | | | | |
Db 445 PCQASGKPTPGISWLRDGLPIDITDSRISQHSSTGLHIADLKKPDGTGYTCIAKNEDGES 504

Qy 176 SWSAVLDVTE--SGATISKNYDLSDLPGPSKPQVTDVTKNSVTLSWQ-PGTPGTLPSA 232
| | | | | | | | | | | | | | | | | | | | | |
Db 505 TWSASLTVEDHTSNAQFVRMPDPSPSSPTQPIIVNVTDTEVELHWNAPSTSGAGPITG 564

Qy 233 YIIIEAFSQSVNSWQTVANHVKTLYTVRGLRPNTIYLFMVRIN-----PKVS---VT 283
| | | | | | | | | | | | | | | | | | | | | |
Db 565 YIIQYSPDLGQWFNIPDVASTEYRIKGLKPSHSYMFVIRAEKIGITPSVSSALVT 624

Qy 284 QXKP 287
| |
Db 625 TSKP 628

RESULT 5
T29548
hypothetical protein: ZK377.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29548
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid ZK377.
A:Reference number: Z20639
A:Accession: T29548
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-874 <NHA>
A:Cross-references: EMBL:U88183; PIDN:AAB52657.1; GSPDB:GN00028; CESP:ZK377.2
A:Experimental source: strain Bristol N2; clone ZK377
C:Genetics:
A:Gene: CESP:ZK377.2
A:Map position: X
A:Introns: 91/2; 356/1; 452/1; 701/3; 746/3; 850/1

Query Match 17.0%; Score 388; DB 2; Length 874;
Best Local Similarity 38.8%; Pred. No. 9.4e-16;
Matches 83; Conservative 38; Mismatches 81; Indels 12; Gaps 4;

Qy 86 VTDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDPLPVISWLKEGFTFPGDRPRATI 145
| | | | | | | | | | | | | | | | | | | | | |
Db 20 VTGNTPAKPPPTIEHGHQNTLMVGSAILPCQASGKPTPGISWLRDGLPIDITDSRISQ 79

Qy 146 EQGTILQIKNLRIISDTGTYTCVATSSSGEASWSAVLDVTE--SGATISKNYDLSDLPGPS 203
| | | | | | | | | | | | | | | | | | | | | |
Db 80 HSTGSLHIADLKKPDGTGYTCIAKNEDGESTWSASLTVEDHTSNAQFVRMPDPSPFPSSP 139

Qy 204 SKPQVTDVTKNSVTLSWQ-PGTPGTLPSAYIIIEAFSQSVNSWQTVANHVKTLYTVRG 262
| | | | | | | | | | | | | | | | | | | | | |
Db 140 TQPIIVNVTDTEVELHWNAPSTSGAGPITGYIIQYSPDLGQWFNIPDVASTEYRIK 199

```

RESULT 7
S50893
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type sigma precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jan-2000
C:Accession: S50893; S40281
R:Wagner, J.; Boerboom, D.; Tremblay, M.L.
Eur. J. Biochem. 226, 773-782, 1994
A:Title: Molecular cloning and tissue-specific RNA processing of a murine receptor-type
A:Reference number: S50893; MUID:95112841
A:Accession: S50893
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1907 <WAG>
A:Cross-references: EMBL:X82288; NID:a587483; PIDN:CAA57732.1; PID:a587484

```

```

Qy      1 QIVAQGRTVTFPCCEITKGNPQPAVFVWKEQGSQNLFFPNQPPQNSRCSVSPGTDLTITNIQ 60
      :: | : | | : | | | : | : | : | : | :
Db    144 KVERTRTATMLCAASGNPDPEITWFKD----FLPVPDSASNGRIQLRSGALQIESSE 198

Qy     61 RSDAGYYICQALTVAG---SILAKAQLEVDVLTDRPPPIILQGPNQTLAVDGTALLKC 117
      : | | | | : | | : | : | : | : | : | :
Db    199 ETDQKGYECVATNSAGVRYSSPANLYVRVRV---APRFSIL--PMSHEIMPGGNVNITC 253

Qy    118 KATGDPLPVISWLKEGFTFPPGRDPRTATIEQGEQTLQIKNLRISDTGTYTCVATSSSGEASW 177
      | | | : | : | : | : | : | : | : | : | : | :
Db    254 VAVGSPMPYVQWKMQGAEDLTPEDDMPV---GRNVLELTDVKDSANYTCVAMSLG--- 305

Qy    178 SAVLDVTESGATISKNYDLSDLPGPPSPKQVDTVTKNSVLTSWQPGTPTGLPASAYIIEA 237
      | | : | : | : | : | | | : | : | : | : | : | : | :
Db    306 -----VIEAVAQIT---VKSLPKAPGTPVVTIENTATSITVTWDSGNPD--PVSYYIVIEY 354

Qy    238 FSQSVSNWSQTVANHVKTLLTYVRGLRPNTIYLFMVRAINPKVSVTQXKPQRKNNGS-TWA 296
      | : | : | : | : | : | | | : | : | : | : | : | :
Db    355 KSKSQDGPYQ- IKEDITTTTRYSIGGLSPNSEYEIWWSAVN---SIGQPPSPESVVTRTGE 410

Qy    297 NVPLPPP-----PVQP-----LPGTELEHYAVEQQENGYDSDSWCPPL 334
      | | | | | : | : | : | : | : | : | : | : | :
Db    411 QAPASAPRNVQARMLSATTTMIVQWEEPVEPENGRLIRYRV-IYIME-----PEH 457

Qy    335 PVQTYLHQGLED-----ELEEDD-----DRVPTPPVRGVASSP-- 367
      | | : | : | : | : | : | : | : | : | : | : | :
Db    458 PVGNWQKHNVDSLLTTVGSLLEDETYTVRVLAFTSVGDGPLSDPIQVTKQGVPGQPMN 517

Qy    368 -----AISF-----GQQSTATITLPSPR---EEMQP---- 389
      | : | : | : | : | : | : | : | : | : | : | :
Db    518 LRAEAKSETSIGLSWSAPRQESVIKYELLFREGDRGREVGRTFDPTTAFVVEDLKPNTEY 577

Qy    390 ---MLQASP---XFTS-----SQRPRPTSP---FSTDNNTSAAQSQRPRPTKKHKG 433
      | : | : | : | : | : | : | : | : | : | : | :
Db    578 AFLAARSPOGLGFTAVVHORTLOAKPSAPPDQVKCTSLRSTAILVSWRPPPPETHNG 636

```

```

RESULT      8
I50212
protein-tyrosine-phosphatase (EC 3.1.3.48) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jan-2000
C;Accession: I50212
R;Stoker, A.W.
Mech. Dev. 46, 201-217, 1994
A;Title: Isoforms of a novel cell adhesion molecule-like protein tyrosine phosphatase

```

RESULT 10
S46217
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat
N;Alternate names: leukocyte common antigen-related phosphatase

```

Qy      1 QIVAQGRITVFPCTKGNPQAPVFWQKEGSQLLPNQPPQPNRSKVSVPSTGDLTITNIQ 60
      ::|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    144 KVVERTRTATMLCAASGNPDPEITWFKD----FLPVDPSASNGRIKQLRSGALQIESSE 198

Qy     61 RSDAGYYICQALTVAG---SILAKAQLEVTDVLTDRPPPIILQGPNQTLAVDGTALLKC 117
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    199 ETDQGGKYECVATNSAGVRYSSPANLYVRVRV---APRFSIL--PMSHEIMPGGMNVNITC 253

Qy    118 KATGDPLPVISWLKEGFTFPGRPDRATIQEQTGLQIKNLRISDGTGTYTCVATSSSGEASW 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    254 VAVGSPMPYVKMKGAEGLTPEDDMPV---GRNVLELTDVKDSANYTCVAMSSLG---- 305

Qy    178 SAVLDVTESGATISKNYLSDLPGPPSKPQVTDVTKNSVTLSWQPGTPTGPLASAYIIIEA 237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    306 ----VIEAVAQIT---VKSLPKAPTPTVVNTATSTITVWDSGNPD--PVSYYIVIEY 354

Qy    238 FSQSVSNSWQTVANHVKTTLYTVRGLRPNTIYLFMVRAINPKVSVTQXKPQKNNGS-TWA 296
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    355 KSKSQDGPYQ-IKEDITTTTTSIGGLSPNSEYEIWWSAVN---SIGQPPSPESVVTRTGE 410

Qy    297 NVPLPPP-----PVQP---LPGTELEHYAVEQQENGYSDSWCPPL 334
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      411 QAPASAPRNVQARMLSATTMTIVQWEEPEVPEPGLIRGYRV-YYTME-----PEH 457
Qy      335 PVQTYLHQGLE-----ELEED-----DRVPTPPVRGVASSP-- 367
      || : :|| | ||: | : || |
Db      458 PVGNWQKHNVDDSLLTTVGSLLEDETYTVRVLAFTSVGDGPLSDPIQVKTQQGVPGQPMN 517
Qy      368 --AISFGQQTSTATLTSPSREEMQMLQASPXFTSSQRPR-----PTSPFSDS---NT 415
      : : : | :||: | :|| : || : ||
Db      518 LRAEAKSETSIGLSWAPRQE--SVIKYELLFREGDRGREGVGRTFDPTAFVVEDLKPT 575
Qy      416 SAALSQSQR 424
      : : |
Db      576 EYAFRLAAR 584

```

RESULT 11
 C54689
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form B precursor
 N;Alternate names: MPTP delta type B/C
 N;Contains: protein-tyrosine phosphatase, receptor type delta, splice form C
 C;Species: Mus musculus (house mouse)
 C;Date: 25-Apr-1995;sequence_revision 19-May-1995 #text_change 12-Feb-1999
 C;Accession: C54689; B54689
 R;Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.
 Mol. Cell. Biol. 13, 5513-5523, 1993
 A;Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in special
 A;Reference number: A54689; MUID:93360986
 A;Accession: C54689
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-1894 <MIZ>
 A;Experimental source: brain; splice form B
 A;Note: sequence inconsistent with nucleotide translation
 A;Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBIP:137487)
 A;Accession: B54689
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-352,'H',354-535,'S',537-601,1002-1894 <MI2>
 A;Experimental source: brain; splice form C
 A;Note: sequence inconsistent with nucleotide translation
 A;Note: sequence extracted from NCBI backbone (NCBIN:136527, NCBIP:136530)
 C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology
 Cg
 C;Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester
 F;45-107/Domain: immunoglobulin homology <IMM1>
 F;245-299/Domain: immunoglobulin homology <IMM2>
 F;317-399/Domain: fibronectin type III repeat homology <FN3A>
 F;1278-1894/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F;1652-1874/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F;1536/Active site: Cys (phosphocysteine intermediate) #status predicted
 F;1542/Binding site: substrate phosphate (Arg) #status predicted
 F;1826/Active site: Cys (phosphocysteine intermediate) #status predicted
 F;1832/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.9%; Score 317; DB 2; Length 1894;
Best Local Similarity 23.4%; Pred. No. 4e-11;
Matches 128; Conservative 63; Mismatches 188; Indels 168; Gaps 22;

[illegible]

```

Db      305 -----VIEAIAQIT---VKALPKPGTPVVVTESTATSTLTWDSGNPG--PVSYYIIQ 352
Qy      237 AFSQSVSNWQTVANHVKTTLYTVGRLRPNTIYLEFMVRAIN----- 277
      ::      :: :: :: :: :: :: :: :: :: :: :: ::
Db      353 EKPKNSEEPYKEI-DGIATTRYISVAGLSPYSDYEFRVVAVNNIGRGPASEPVLTTQTSQT 411
Qy      278 ---PK-----VSVTQXKPQKNNG-----STWAN----- 297
      |:      |:      |:      |:      |:      |:      |:      |:
Db      412 PSSAPRDVQARMLSSTTILVQWKPEEPNGQIQGYRVYYTMDPTQHVNMMKHNVDASI 471
Qy      298 -----VPLPPPPVQPL-----PGTELEHYAVEQQENGYSDSDS 329
      ||      ||      ||      ||      ||      ||      ||      ||
Db      472 TTIGNLVPQKTYISVKVLAFTSIGDGPLSSDIQVITQTGPGQPLNFKAEPESETSI-LLS 530
Qy      330 WCPPL--PVQTYLHGQLEDELEDDRVPTPPVRGVASSPAISFGQOSTATLTPTSPREM 387
      | |      : |      : |      : |      : |      : |      : |
Db      531 WTTPREDTIASY-----ELWRDGDQ-----GEEQRITIEPGTSYRL 567
Qy      388 QPM-----LQASPYXTSS---QRP-RPTSPFSTSDNTSAAALSQSQRPR 426
      | :      | :      | :      | :      | :      | :      | :
Db      568 GQLKPNLSYIFRLSATSPQGLASTAEISARTMQKPSPPQDISCTSPSSTLSILVSWQPP 627
Qy      427 PTKKHKG 433
      | :      | :
Db      628 PVEKQNG 634

```

RESULT 12

ID:DHULK
 leukocyte antigen-related protein precursor - human
 N;Alternate names: leukocyte common antigen homolog
 N;Contains: protein-tyrosine-phosphatase (EC 3.1.3.48)
 C;Species: Homo sapiens (man)
 C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 22-Jun-1999
 C;Accession: S03841; JL0051
 R;Streull, M.; Krueger, N.X.; Hall, L.R.; Schlossman, S.F.; Saito, H.
 J. Exp. Med. 168, 1523-1530, 1988
 A;Title: A new member of the immunoglobulin superfamily that has a cytoplasmic region ho
 A;Reference number: JL0051; MUID:89035978
 A;Accession: S03841
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1897 <STR>
 A;Cross-references: EMBL:Y00815; NID:g34266; PIDN:CAA68754.1; PID:g34267
 C;Genetics:
 A;Gene: GDB:PTPRF; LAR
 Cross-references: GDB:120138; OMIM:179590
 Map position: lp34-lp34
 C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;
 ogy
 C;Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
 F;1-16/Domain: signal sequence #status predicted <SIG>
 F;17-1897/Product: leukocyte antigen-related protein #status predicted <MAT>
 F;17-1250/Domain: extracellular #status predicted <EXT>
 F;37-99/Domain: immunoglobulin homology <IMM1>
 F;139-199/Domain: immunoglobulin homology <IMM2>
 F;236-290/Domain: immunoglobulin homology <IMM3>
 F;308-390/Domain: fibronectin type III repeat homology <FN3A>
 F;403-489/Domain: fibronectin type III repeat homology <FN3B>
 F;501-583/Domain: fibronectin type III repeat homology <FN3C>
 F;596-685/Domain: fibronectin type III repeat homology <FN3D>
 F;698-798/Domain: fibronectin type III repeat homology #status atypical <FN3E>
 F;810-893/Domain: fibronectin type III repeat homology <FN3F>
 F;905-989/Domain: fibronectin type III repeat homology <FN3G>
 F;1001-1078/Domain: fibronectin type III repeat homology <FN3H>
 F;1251-1274/Domain: transmembrane #status predicted <TMM>
 F;1275-1897/Domain: intracellular #status predicted <INT>
 F;1285-1897/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F;1365-1586/Domain: protein-tyrosine-phosphatase homology <PTP1>
 F;1654-1877/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F;44-97,146-197,243-288/Disulfide bonds: #status predicted
 F;107,240,285,711,956/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;1538/Active site: Cys (phosphocysteine intermediate) #status predicted

F:1544/Binding site: substrate phosphate (Arg) #status predicted
F:1829/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1835/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.5%; Score 308.5; DB 1; Length 1897;
Best Local Similarity 28.7%; Pred. No. 1.3e-10;
Matches 118; Conservative 47; Mismatches 165; Indels 81; Gaps 17;

```
Qy      1 QIVAQGRVTYTFPCETKGNPQPAVFVWKEGSQLNLFPPNQPPQNRSRCSVSPTGLDTITNIQ 60  
       :|::||| | ||| ::|:: | | | :|::|:  
Db     134 KVVEKARTATMLCAAGGNPDPEISWFKD-----FLPVPDPATSNGRIKQLRSGALQIESSE 188  
  
Qy      61 RSDAGYYICQALTAVGS-ILAKAQLVTDVLTRDRPPIILQGPNQTLAVDGTALLCKKA 119  
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     189 ESDQGYECVATNSAGRTRYSAANLVR--VRRVAAPRFSSIPSSGEVMPGGSVNLTCVA 245  
  
Qy     120 TGDPLPVISWL-----KEGTFPGRPDRATIQEQGTLIKNNLRISDRTGTYTCVATSSS 172  
       |:|:|:| | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     246 VGAPMPYVKMMGAELTKEDEMPVGRN-----VLELSN--VVRSANYTCVAISSL 294  
  
Qy     173 GEASWSAVLDVTESGATISKNYDLSDLGPSPSKPQVTDTKNSVLTSLWGPTGCTLPASA 232  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     295 G-----MIEATAQVT---VKALKPPIDLVTETTATSVTLTWDSG--NSEPVYI 339  
  
Qy     233 YIIIEAFSQSVSNSWQTVANHVKTLYTVRGLRPNTIYLFWRAINPKVSVTXKKPQKN-N 291  
       | | | : | | | : | | | | | | | | | | | | | | | | | | | | | |  
Db     340 YGIQYRAACTEGFPQEVDGVATTTRYSIGGLSPFSEYAFRVLAVN---STIGCGPPSEAVR 395  
  
Qy     292 GSTWANVPLPPP-VQPQ---LPGTELEHYAVEQQENG-----YDSDSWCPP----- 333  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     396 ARTGEQAPSPPRRVQARMSASTMLVQWEPPPEENGLVRGYRYVYTSDSRPPNAWHKH 455  
  
Qy     334 -----LPVQTYLHQGLEDELEEDDRVPVTPVR---GVASSPA 368  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     456 NTDAGLLTTVGSLLPGITYSRLRVLAFTAVDGGPSPTIQVKTQOGVPAQA 506
```

RESULT 13

A56178
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - human
N;Alternate names: protein-tyrosine-phosphatase BTPP-2
C;Species: Homo sapiens (man)
C;Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 21-Jan-2000
C;Accession: A56178; S12052; B44929
R;Pulido, R.; Krueger, N.X.; Serra-Pages, C.; Saito, H.; Streuli, M.
J. Biol. Chem. 270, 6722-6728, 1995
A;Title: Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta isoforms. /
A;Reference number: A56178; MUID:95204468
A;Accession: A56178
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1912 <PUL>
A;Cross-references: GB:L38929; NID:g755652; PIDN:AAC41749.1; PID:g755653
R;Krueger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990
A;Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatase
A;Reference number: S12049; MUID:91006018
A;Accession: S12052
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 390-1912 <KRU>
A;Cross-references: GB:X54133; NID:g35789; PIDN:CAA38068.1; PID:g35790
A;Note: the sequence from Fig. 5B is inconsistent with that from Fig. 5A in having 56
R;Adachi, M.; Sekiya, M.; Arimura, Y.; Takekawa, M.; Itoh, F.; Hinoda, Y.; Imai, K.;
Cancer Res. 52, 737-740, 1992
A;Title: Protein-tyrosine phosphatase expression in pre-B cell NALM-6.
A;Reference number: A44929; MUID:92119637
A;Accession: B44929
A;Molecule type: mRNA
A;Residues: 1756-1804,'C',1806-1845 <ADA>
A;Cross-references: GB:S78086; NID:q243545; PIDN:AAB21147.1; PID:q243546

A;Experimental source: pre-B cell NALM-6
A;Note: sequence extracted from NCBI backbone (NCBIN:78086, NCBIP:78087)
A;Note: the authors did not report the entire codon for residue 90
C;Genetics:
A;Gene: GDB:PTPRD
A;Cross-references: GDB:131384; OMIM:601598
A;Map position: 9p24-9p24
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;
ogy
C;Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembrane
F;38-100/Domain: immunoglobulin homology <IMM1>
F;140-209/Domain: immunoglobulin homology <IMM2>
F;250-304/Domain: immunoglobulin homology <IMM3>
F;711-811/Domain: fibronectin type III repeat homology <3FR>
F;1293-1912/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;1669-1892/Domain: protein-tyrosine-phosphatase homology <PTP2>
F;1553/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1559/Binding site: substrate phosphate (Arg) #status predicted
F;1844/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1850/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.5%; Score 307.5; DB 2; Length 1912;
Best Local Similarity 24.1%; Pred. No. 1.5e-10;
Matches 130; Conservative 64; Mismatches 205; Indels 141; Gaps 22;

Qy 1 QIVAGRTVTTPCETKGNPQPAVFWQKEGSQLNLPNQPPQNSRC-----SVSPST--- 51
Db 135 KVVERTRTATMLCAASGNPDPEITWFKD-----FLPVDTSNNNGRIKQLRSEISGGTPIR 189
Qy 52 GDLITNIQRSDAGYICQALTAVGS-ILAKAQLEVTDLTDRPPPIILQGPANQTLAV 109
Db 190 GALQIEQSEESDQGYECVATNSAGTRYAPANLVKREVRVPRFSPISPPNHEIMP 249
Qy 110 DGTALLKCKATGDPPLVSWLKEGFTFPGRDPRATIQEQTILQIKNLRIISDTGTYTCVAT 169
Db 250 GGSVNTICVAVGSPMPYVKWMLGAEDLTPEDDMPI--GRNVLELNDVR--QSANVTCVAM 305
Qy 170 SSSGEASWSAVLDVTESGATISKNYDLSDLPGPSKQPTVDTKNSVTLWSQPGTPTLP 229
Db 306 STLG-----VIEAIAQIT---VKALPKPPGTPTVTESTATSTLWDSGNPE--P 350
Qy 230 ASAYIIIEAFSQSVSNQTVANHVHTTLYTVRGLRPNTIYLFMVRAIN----- 277
Db 351 VSYIYIHKPKNSEELYKEI-DGVATTRYSVAGLSPIYSYEFRRVAVNNIGRGPPEPVL 409
Qy 278 -----PK-----VSVTQXKPKQKNG-----STWAN- 297
Db 410 TQTSEQAPSSAPRDVQARMLSSSTILVQWKEPEEPNGQIQGYVYVYTMPTQHNWMMKH 469
Qy 298 -----VLPPLPPVQPL-----PQTELEHYAVEQQE 322
Db 470 NVADSQITIGNLVQPKTYSVKVLAFSTIGDGLSSDIQVITQTGVPGQLNFAEPESE 529
Qy 323 NGYDSDSWCPPLPVQTYLHQGL--EDELLEDDRVPTTP-----VRGVASSPAISF--GQ 373
Db 530 TSI-LLSWTPPRS-DTIANYELVKDGEHGEORITIEPTGSYRLQGLKPNSLYFRLAA 587
Qy 374 QSTATLTPSPREEMQPLQASPXFTSSQRPRTSPFSTDSNTSAAQSQRPRPTKKHK 433
Db 588 RSPQGLGASTAEISARTMQSKPS-----APPQDISCTSPSTSLVSNQPPVVEKQNG 640

RESULT 14

S46216
leukocyte antigen-related protein precursor - rat
N;Alternate names: leukocyte common antigen homolog
N;Contains: protein-tyrosine-phosphatase (EC 3.1.3.48)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C;Accession: S46216; S23252; A41032; A33154
R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.
Biochem. J. 302, 39-47, 1994

A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-
A;Reference number: S46216; MUID:94347119
A;Accession: S46216
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1898 <ZHA>
A;Cross-references: EMBL:L11586; NID:g205132; PIDN:AAC37655.1; PID:g205133
R;Hashimoto, N.; Zhang, W.R.; Goldstein, B.J.
Biochem. J. 284, 569-576, 1992
A;Title: Insulin receptor and epidermal growth factor receptor dephosphorylation by t
A;Reference number: S23126; MUID:92287069
A;Accession: S23252
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1361-1604;1649-1898 <HAS>
R;Pot, D.A.; Woodford, T.A.; Remboutsika, E.; Haun, R.S.; Dixon, J.E.
J. Biol. Chem. 266, 19688-19696, 1991
A;Title: Cloning, bacterial expression, purification, and characterization of the cyt
A;Reference number: A41032; MUID:92011772
A;Accession: A41032
A;Molecule type: mRNA
A;Residues: 1035-1072,'S',1074-1433,'T',1435-1638,'N',1640-1642,'HT',1645-1898 <POT>
A;Cross-references: GB:M60103; NID:g205130; PIDN:AAA41510.1; PID:g205131
R;Pot, D.A.; Woodford, T.A.; Remboutsika, E.; Haun, R.S.; Dixon, J.E.
submitted to the Protein Sequence Database, December 1990
A;Reference number: A33154
A;Accession: A33154
A;Molecule type: mRNA
A;Residues: 1035-1072,'S',1074-1433,'T',1435-1638,'N',1640-1642,'HT',1645-1898 <PO2>
C;Comment: Only the first of the two domains homologous with protein-tyrosine-phospha
C;Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog
ogy
C;Keywords: duplication; glycoprotein; phosphoprotein; phosphoric monoester hydrolase
F;1-27/Domain: (or 1-26) signal sequence #status predicted <SIG>
F;28-1898/Product: (or 27-1898) leukocyte antigen-related protein #status predicted
F;28-1251/Domain: (or 27-1251) extracellular #status predicted <EXT>
F;47-109/Domain: immunoglobulin homology <IMM1>
F;149-209/Domain: immunoglobulin homology <IMM2>
F;246-300/Domain: immunoglobulin homology <IMM3>
F;318-400/Domain: fibronectin type III repeat homology <FN3A>
F;413-499/Domain: fibronectin type III repeat homology <FN3B>
F;511-593/Domain: fibronectin type III repeat homology <FN3C>
F;606-695/Domain: fibronectin type III repeat homology <FN3D>
F;708-799/Domain: fibronectin type III repeat homology <FN3E>
F;811-895/Domain: fibronectin type III repeat homology <FN3F>
F;906-990/Domain: fibronectin type III repeat homology <FN3G>
F;1002-1079/Domain: fibronectin type III repeat homology <FN3H>
F;1252-1275/Domain: (or 1259-1275) transmembrane #status predicted <TM>
F;1276-1898/Domain: intracellular #status predicted <INT>
F;1286-1898/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F;1366-1587/Domain: protein-tyrosine-phosphatase homology <PTP1>
F;1655-1878/Domain: protein-tyrosine-phosphatase homology <PTP2>
F;54-107,156-207,253-298/Disulfide bonds: #status predicted
F;117,250,295,721,957/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1539/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1545/Binding site: substrate phosphate (Arg) #status predicted
F;1830/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1836/Binding site: substrate phosphate (Arg) #status predicted

Query Match 13.4%; Score 305.5; DB 2; Length 1898;
Best Local Similarity 25.0%; Pred. No. 1.9e-10;
Matches 134; Conservative 58; Mismatches 200; Indels 143; Gaps 24;

Qy 1 QIVAGRTVTTPCETKGNPQPAVFWQKEGSQLNLPNQPPQNSRCVSPTGDLITNIQ 60
Db 144 KVEKARTATMLCAAGGNPDPEISWFKD-----FLPVDPASSNGRIKQLRSGALQIESSE 198
Qy 61 RSDAGYICQALTAVGS-ILAKAQLEVTDLTDRPPPIILQGPANQTLAVDGTALLCKKA 119
Db 199 ESDQGYECVATNSAGTRYAPANLVLR---VRRVAPRFSIPPSSQEVMPGNNVLTCA 255
Qy 120 TGDPLPVISWL-----KEGFTFPGRDPRATIQEQTILQIKNLRIISDTGTYTCVATSSS 172

Db 256 VGAPMPYKVMGAEELTKEDMPVGRN-----VLELSN--VMRSANTCVAISSL 304
Qy 173 GEASWSAVLDVTESGATISKNYLSDLPGPSKQVTDVTKNSVTLWSQPGTLPASA 232
Db 305 G-----MIEATAQVT-----VKALPKPPIDLVTETATSVTLTWDG--NTEPVSF 349
Qy 233 YIEAFSQSVSNWQTVANHVKTTLTYVRGLRPNTIYLFMRVAINPKVSVTQXKPQN-N 291
Db 350 YGIQYRAAGTDGPFQEV-DGVASTRYSIGLSPFSEYAFRLAVN---SIGRPPSEAVR 405
Qy 292 GSTWANVPLPPP-PVQP---LPGTELEHYAVEQQENG-----YDSDS-----W--- 330
Db 406 ARTGEQAPSSPPRRVQARMLASATMLVQWPEEPNGLVRGYRVYITPDSRRPLSAWHKH 465
Qy 331 -----CPPLP-VQTYLHQGLEDE-----LEEDD 352
Db 466 NTDAGLLTVGSLLPGITYSRLVLAFTAVGDGPPSPTIQVKTQQGVPAQPADFQAKAESD 525
Qy 353 DRVP---TPPVRGVASSPAISF-----GQQSTATLTPSPR---EEMQP----- 389
Db 526 TRIQLSWLLPPQERIIKYLWYAAEDEGQHKVTFDPTISSYTLLEDLKPDTLYHFQLAAR 585
Qy 390 -----MLQASPXFTSSQRPRTSPFSDSNTSALSQSQRPRPTKKHK 433
Db 586 SDLGVGVFTPTVEACTAQSTSPAPKQKTCVSTGTT---VRVSWVPPPADSRNG 637

RESULT 15

T43027

neural cell adhesion molecule L1 - goldfish

N;Alternate names: E587 antigen

C;Species: Carassius auratus (goldfish)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000

C;Accession: T43027

R;Giordano, S.; Laessing, U.; Lottspeich, F.; Stuermer, C.A.O.

submitted to the EMBL Data Library, April 1996

A;Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule exp

A;Reference number: Z22294

A;Accession: T43027

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1232 <GIO>

A;Cross-references: EMBL:U55211; NID:g1305526; PID:g1305527; PIDN:AAA99159.1

C;Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; 1

C;Keywords: cell adhesion; membrane protein

Query Match 13.0%; Score 297.5; DB 2; Length 1232;
Best Local Similarity 24.9%; Pred. No. 3.4e-10;
Matches 111; Conservative 67; Mismatches 184; Indels 83; Gaps 17;

Qy 5 QGRIVTFPCETKGNPQPAVFWQKEGSQLNLFNPQPPQNSRCSVSPGDLTITNIQRSDA 64
Db 409 EGQTVLLQCRFTFGSPQPKVDWQITNS-----GPALANAKMSQTSQNLQISDVSEEDS 461
Qy 65 GYICQALTAVGASILAKAQLEVTDLVTRPPPI-ILQGPANQTLAVDGTALLKCKATGDP 123
Db 462 SMYTCVSTSNMSISAEVLVLRNRTKIVDPQDLRLVLRG-----DDAVLQCRYTVDH 512
Qy 124 L---PVISWLKE--GFTFPGRPRTIQEQGLQIKNLRIQSDTGYTCVATSSSGEASWS 178
Db 513 MLKQPPIQWKKDKHKITSSANDDKYTESPDGSLKITDVQMEDSGIYSC-----EIS 563
Qy 179 AVL-D-VTESGATISKNYLSDLPGPSKQVTDVTKNSVTLWSQPGTLPASAYIEA 237
Db 564 TKLDSVSATGSIV-----VLDKPGSPHSLSEKKERSVTLWSMGAENNSPISEYVIER 618
Qy 238 FSQS--VSNWQTVANHVKTTLTYVRGLRPNTIYLFMRVAINPKVSVTQXKPQKNGSTW 295
Db 619 KEKQNPQKQWEEYRRVPQDITHLEIHLPYSTYHFRVRGVN-GIGMSESPSPSESYST- 676
Qy 296 ANVPLPPPPVQP-----LPGTELEHYAVEQQENGSDSWCPPLPVQTYLH 341

Db 677 ---PAAKPDMNPENVTSTVSDNSLVITWQLE----QRQFNG-----PGFKYKIYWR 722
Qy 342 QGLEDELEEDDDRVPPTPPVRGVASSPAISFGQQSTATLTPSPREEMQ-----PMLQA 393
Db 723 QEGDSHWMESSASNPFFIVEGPGTFIPRQIKVQAVNELGAGPEPDAGIGYSGEDLP-LEA 781
Qy 394 SPXFTSSQRPRT-----SPFSTDS 413
Db 782 PSSVAVSELNKTTLVKWSPVSTKS 806

Search completed: January 22, 2001, 12:27:28

Job time: 2125 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:29:42 ; Search time 162.41 Seconds
(without alignments)
86.298 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 QIVAQGRVTTPCETKGNPQ.....TSAALSQSQRPRPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309.5	13.6	837	1 NCM2_HUMAN	O15394 homo sapien
2	308.5	13.5	1897	1 PTPF_HUMAN	P10586 homo sapien
3	307.5	13.5	1912	1 PTPD_HUMAN	P23468 homo sapien
4	295.5	13.0	2029	1 LAR_DROME	P16621 drosophila
5	292.5	12.8	837	1 NCM2_MOUSE	O35136 mus musculus
6	284.5	12.5	1377	1 NEO1_RAT	P97603 rattus norv
7	282	12.4	1443	1 NEO1_CHICK	Q90610 gallus gall
8	279	12.2	1493	1 NEO1_MOUSE	P97798 mus musculus
9	268	11.8	1257	1 CAML_HUMAN	P32004 homo sapien
10	267.5	11.7	1461	1 NEO1_HUMAN	Q92859 homo sapien
11	265.5	11.6	1260	1 CAML_MOUSE	P11627 mus musculus
12	261.5	11.5	1259	1 CAML_RAT	Q05695 rattus norv
13	259	11.4	1447	1 DCC_MOUSE	P70211 mus musculus
14	258	11.3	1239	1 NRG_DROME	P20241 drosophila
15	256	11.2	1040	1 AXO1_HUMAN	Q02246 homo sapien
16	251.5	11.0	1036	1 AXO1_CHICK	P28685 gallus gall
17	251	11.0	761	1 NCA2_HUMAN	P13592 homo sapien
18	250.5	11.0	1284	1 NRCA_CHICK	P35331 gallus gall
19	249	10.9	1447	1 DCC_HUMAN	P43146 homo sapien
20	248	10.9	725	1 NCA2_MOUSE	P13594 mus musculus
21	247.5	10.9	1018	1 COMT_HUMAN	Q12860 homo sapien
22	246.5	10.8	1115	1 NCA1_MOUSE	P13595 mus musculus
23	245	10.7	1040	1 AXO1_RAT	P22063 rattus norv
24	244	10.7	1010	1 COMT_CHICK	P14781 gallus gall
25	243	10.7	2012	1 DSCA_HUMAN	O60469 homo sapien
26	241.5	10.6	1020	1 COMT_MOUSE	P12960 mus musculus
27	238.5	10.5	853	1 NCA1_BOVIN	P31836 bos taurus
28	238.5	10.5	858	1 NCA1_RAT	P13596 rattus norv
29	236.5	10.4	848	1 NCA1_HUMAN	P13591 homo sapien
30	235.5	10.3	1906	1 KMLS_CHICK	P11799 gallus gall
31	223	9.8	1266	1 NGCA_CHICK	Q03696 gallus gall
32	222.5	9.8	898	1 FAS2_SCHAM	P22648 schistocerc
33	222.5	9.8	1142	1 MYPF_HUMAN	Q14324 homo sapien

34	222	9.7	1913	1 KMLS_HUMAN	Q15746 homo sapien
35	220	9.6	1051	1 PTK7_CHICK	Q91048 gallus gall
36	218	9.6	811	1 FS22_DROME	P34083 drosophila
37	218	9.6	873	1 FS21_DROME	P34082 drosophila
38	215	9.4	1091	1 NCA1_CHICK	P13590 gallus gall
39	215	9.4	4393	1 PGBM_HUMAN	P98160 homo sapien
40	214.5	9.4	1070	1 PTK7_HUMAN	Q13308 homo sapien
41	213.5	9.4	819	1 FGRI_CHICK	P21804 gallus gall
42	210.5	9.2	822	1 FGRI_HUMAN	P11362 homo sapien
43	210.5	9.2	822	1 FGRI_MOUSE	P16092 mus musculus
44	210.5	9.2	822	1 FGRI_RAT	Q04589 rattus norv
45	210.5	9.2	3707	1 PGBM_MOUSE	Q05793 mus musculus

ALIGNMENTS

RESULT 1
NCM2_HUMAN
ID NCM2_HUMAN STANDARD; PRI; 837 AA.
AC O15394;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE 2 PRECURSOR (N-CAM 2).
GN NCAM2 OR NCAM21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97369930; PubMed=9226371;
RA Paoloni-Giacobino A., Chen H., Antonarakis S.E.;
RT "Cloning of a novel human neural cell adhesion molecule gene (NCAM2)
RT that maps to chromosome region 21q21 and is potentially involved in
RT Down syndrome.";
RL Genomics 43:43-51(1997).
CC -1- FUNCTION: MAY PLAY IMPORTANT ROLES IN SELECTIVE FASCICULATION AND
CC ZONE-TO-ZONE PROJECTION OF THE PRIMARY OLFACTORY AXONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED MOST STRONGLY IN ADULT AND FETAL
CC BRAIN.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75330; AAB80803.1; -.
DR MIM; 602040; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; ig; 5.
KW Cell adhesion; Transmembrane; Glycoprotein; Repeat;
KW Immunoglobulin domain; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 837 NEURAL CELL ADHESION MOLECULE 2.
FT DOMAIN 20 697 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 698 718 POTENTIAL.
FT DOMAIN 719 837 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 100 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 129 193 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 225 288 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 315 387 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 415 481 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 482 581 FIBRONECTIN TYPE-III.

Query Match 13.5%; Score 308.5; DB 1; Length 1897;
Best Local Similarity 28.7%; Pred. No. 3.4e-11;
Matches 118; Conservative 47; Mismatches 165; Indels 81; Gaps 17;

Qy 1 QIVAQGRVTTPCETKGNPQPAVFWQKEGSQLLPNQPPQPPNSRCSVSTGDLTITNIQ 60
Db 134 KVVKEARTATMLCAAGGNPDPEISWFKD-----FLPVDTPATNGRIKQLRSGALQIESSE 188
Qy 61 RSDAGYYICQALTAVGS-ILAKAQLVTDVLTDRPPPIILGQPANQTLAVDGTALLKCKA 119
Db 189 ESDQGYECVATNSAGTRYSPANLVVR---VRRVAPRFSIPSSQEVMPGGSVNLTCTVA 245
Qy 120 TGDPLPVISWL-----KEGFTFPGDRPRATIEQSTLQIKNLRISETGYTTCVATSSS 172
Db 246 VGAPMPYKVMGAEELTKEDMPVGRN-----VLELSN--VVRANSNTCAISSL 294
Qy 173 GEASWSAVLDVTESGATISKNYDLSLPGPPSKQVTDVTKNSVTLQSWQPGTGPLPASA 232
Db 295 G-----MIEATAQVT---VKALPKPIDLVITETATSVTLWDG--NSEPTVY 339
Qy 233 YIIIEAFSQSVSNQWTVANHVKTLLTYVRLRPNTIYLFMVRINPKVSTQXKPKQN-N 291
Db 340 YGIQYRAAGTEGPEDEV-DGVATTRYISIGLSPFSEYAFRLAVN---SIGRPPSEAVR 395
Qy 292 GSTWANVLPPLP-PVQP---LPGTELEHYAVEQENG-----YDSQSWCPP----- 333
Db 396 ARTGEQAPSSPPRRVQARMLASTMLVQWPEEPENGLVGRVRYVYTPDSRRPPNAWHKH 455
Qy 334 -----LPVQTYLHQGLEDELEDDRRVTPPVV---GVASSPA 368
Db 456 NTDAGLLITVGSLLGITYSLRLVLAFTAVDGGPSPSTIQVKTQGGVPAQA 506

RESULT 3

PTPD_HUMAN

ID PTPD_HUMAN STANDARD; PRT; 1912 AA.
AC P23468;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE DELTA PRECURSOR (EC 3.1.3.48) (R-PTP-
DE DELTA).
GN PTPRD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
SEQUENCE FROM N.A.
RX MEDLINE=95204468; PubMed=7896816;
RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT "Molecular characterization of the human transmembrane
RT protein-tyrosine phosphatase delta. Evidence for tissue-specific
RT expression of alternative human transmembrane protein-tyrosine
RT phosphatase delta isoforms."
RL J. Biol. Chem. 270:6722-6728(1995).
RN [2]
RP SEQUENCE OF 390-1912 FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases."
RL EMBO J. 9:3241-3252(1990).
CC -1 CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1 ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC TISSUES DUE TO ALTERNATIVE SPLICING.
CC -1 PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC FROM THE TRANSMEMBRANE SEGMENT.
CC -1 SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1 SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1 SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L38929; AAC41749.1; -.
DR EMBL; X54133; CAA38068.1; -.
DR PIR; S12052; S12052.
DR HSSP; P18052; 1YFO.
DR MIM; 601598; -.
DR INTERPRO; IPR000242; -.
DR INTERPRO; IPR000387; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00102; Y_phosphatase; 2.
DR PFAM; PF00041; fn3; 8.
DR PFAM; PF00047; ig; 3.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00700; PRTYPHPTASE.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Duplication;
KW Immunoglobulin domain; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1912 PROTEIN-TYROSINE PHOSPHATASE DELTA.
FT DOMAIN 21 1265 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1266 1290 POTENTIAL.
FT DOMAIN 1291 1912 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 23 115 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 118 225 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 232 318 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 320 414 FIBRONECTIN TYPE-III.
FT DOMAIN 417 513 FIBRONECTIN TYPE-III.
FT DOMAIN 516 606 FIBRONECTIN TYPE-III.
FT DOMAIN 609 708 FIBRONECTIN TYPE-III.
FT DOMAIN 711 822 FIBRONECTIN TYPE-III.
FT DOMAIN 825 916 FIBRONECTIN TYPE-III.
FT DOMAIN 918 1017 FIBRONECTIN TYPE-III.
FT DOMAIN 1020 1137 FIBRONECTIN TYPE-III.
FT DOMAIN 1375 1618 PROTEIN-TYROSINE PHOSPHATASE.
FT DOMAIN 1619 1912 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 1553 1553 BY SIMILARITY.
FT ACT_SITE 1844 1844 BY SIMILARITY.
FT SITE 1175 1178 CLEAVAGE (POTENTIAL).
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 832 832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 181 189 MISSING (IN KIDNEY ISOFORM).
FT VARSPLIC 226 229 MISSING (IN KIDNEY ISOFORM).
FT VARSPLIC 775 783 MISSING (IN KIDNEY ISOFORM).
FT VARSPLIC 609 1137 MISSING (IN FETAL BRAIN ISOFORM).
FT MUTAGEN 1178 1178 R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.
SQ SEQUENCE 1912 AA; 214759 MW; 3AE8CBDC32182E26 CRC64;

Query Match 13.5%; Score 307.5; DB 1; Length 1912;
Best Local Similarity 24.1%; Pred. No. 4e-11;
Matches 130; Conservative 64; Mismatches 205; Indels 141; Gaps 22;

Qy 1 QIVAQGRVTTPCETKGNPQPAVFWQKEGSQLLPNQPPQPPNSRCSVSTGDLTITNIQ 51
Db 135 KVVETRTATMLCAAGGNPDPEITWFKD-----FLPVDTSNNNGRIKQLRSEISIGTPIR 189
Qy 52 GDLTITNIQSDAGYYICQALTAVGS-ILAKAQLVTDVLTDRPPPIILGQPANQTLAV 109
Db 190 GALQIEQSESDQGYECVATNSAGTRYSPANLVRELRRVRRVPPRFSIPTNHEIMP 249

```

Qy 110 DGTALLKCAKATGDPLVISWLKEGFTFPGDRPRATIQEQGTQLQIKNLRIISDTGTYTCVAT 169
   | : : | | | : : | : : | : : | : | | |
Db 250 GGSVNITCAVAGSPMPYVKWMLGAEDLTPEDDMPI--GRNVLELNDVR--QSANYTCVAM 305

Qy 170 SSSGEASVAVLDTVESGATISKNYDLSDLPGPPSKPQVDTVKNSVTLWSQPGTPTGLP 229
   | : | | : | : | : | | | : | : | : | |
Db 306 STLG-----VIEAIAQIT---VKALPKPPGTPVWTESTATISITLTWDSNGPNE--P 350

Qy 230 ASAYIIIEAFSQSVSNQWTVANHVKTTLYTVRGLRPNTIILEMVRAIN----- 277
   | | | : : : : | | : | | | : | : |
Db 351 VSYIYIQHKPKNSEELYKEI-DGVATTRYSVAGLSPYSDYEFRRVAVNNIGRGPPSEPVL 409

Qy 278 -----PK-----VSVTQXPKQKNNG-----STWAN- 297
   | : : | : : | : : | : |
Db 410 TQTSEQAPSSAPRDVQARMLSSTTILVQWKEPEEPNGIQIGRYVYTTMDPTQHVNNWMKH 469

Qy 298 -----VPLPPPPVQPL-----PGTELEHYAVEQQE 322
   | | : | | | | | : |
Db 470 NVADSQITIGNLVPKQYTSYKVLAFSTSIGDGPLSSDIQVITQGVGPQPLNFKAPESE 529

Qy 323 NGYDSDSWCPLPVQTYLHQGL--EDELEDDDRVPTPPP-----VRGVASSPAISF--GQ 373
   | | | | : : | : : : : | : : | : |
Db 530 TSI-LLSWTPPRS-DTIANYELVYKDGEHGEEQRITIEPGTSYRGLGKPNLSLYFRLAA 587

Qy 374 QSTATLTPSPREEMQMLQASPFXTSSQRPRTSPFSDTSNTSAAALSQSQRPRTTKHKHG 433
   : | | | | : | : | : | : | : | : |
Db 588 RSPGGLGASTAIEISARTMOSKPS-----APPDISCTSPSSSTILVSWOPPEYKONG 640

```

RESULT 4

```

LAR_DROME
ID LAR_DROME          STANDARD;      PRT;  2029 AA.
AC P16621;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE DLAR PRECURSOR (EC 3.1.3.48) (PROTEIN-
TYROSINE-PHOSPHATE PHOSPHOHYDROLASE).
GN LAR.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046860; PubMed=2554325;
RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
RT "A family of receptor-linked protein tyrosine phosphatases in humans
and Drosophila.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE=96178473; PubMed=8598047;
RA Krueger N.X., van Vactor D., Wan H.I., Gelbart W.M., Goodman C.S.,
RA Saito H.;
RT "The transmembrane tyrosine phosphatase DLAR controls motor axon
RT guidance in Drosophila.";
RL Cell 84:611-622(1996).
CC -!- FUNCTION: IT IS POSSIBLE THAT DLAR IS A CELL ADHESION RECEPTOR.
CC IT POSSESSES AN INTRINSIC PROTEIN TYROSINE PHOSPHATASE ACTIVITY
CC (PTPASE). IT CONTROLS MOTOR AXON GUIDANCE.
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
CC PIONEER NEURONS IN THE EMBRYO.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M27700; AAA28668.1; -.
DR EMBL; U36857; AAC47002.1; -.
DR EMBL; U36849; AAC47002.1; JOINED.
DR EMBL; U36850; AAC47002.1; JOINED.
DR EMBL; U36851; AAC47002.1; JOINED.
DR EMBL; U36852; AAC47002.1; JOINED.
DR EMBL; U36853; AAC47002.1; JOINED.
DR EMBL; U36854; AAC47002.1; JOINED.
DR EMBL; U36855; AAC47002.1; JOINED.
DR EMBL; U36856; AAC47002.1; JOINED.
DR PIR; A36182; TDFFLK.
DR HSSP; P28827; 1RPM.
DR FLYBASE; FBgn0000464; Lar.
DR INTERPRO; IPR000242; -.
DR INTERPRO; IPR000387; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00102; V_phosphatase; 2.
DR PFAM; PF00041; fn3; 9.
DR PFAM; PF00047; ig; 3.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00700; PRTPHPHASE.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW Cell adhesion; Immunoglobulin domain; Duplication.
FT SIGNAL 1 32
FT CHAIN 33 2029 PROTEIN-TYROSINE PHOSPHATASE DLAR.
FT DOMAIN 33 1377 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1378 1402 POTENTIAL.
FT DOMAIN 1403 2029 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 50 118 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 154 216 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 249 308 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 320 417 FIBRONECTIN TYPE-III.
FT DOMAIN 418 512 FIBRONECTIN TYPE-III.
FT DOMAIN 513 607 FIBRONECTIN TYPE-III.
FT DOMAIN 608 706 FIBRONECTIN TYPE-III.
FT DOMAIN 707 809 FIBRONECTIN TYPE-III.
FT DOMAIN 810 906 FIBRONECTIN TYPE-III.
FT DOMAIN 907 1007 FIBRONECTIN TYPE-III.
FT DOMAIN 1008 1102 FIBRONECTIN TYPE-III.
FT DOMAIN 1103 1207 FIBRONECTIN TYPE-III.
FT DOMAIN 1492 1738 PROTEIN-TYROSINE PHOSPHATASE.
FT DOMAIN 1781 2029 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 1670 1670 BY SIMILARITY.
FT ACT_SITE 1961 1961 BY SIMILARITY.
FT DISULFID 57 111 POTENTIAL.
FT DISULFID 161 209 POTENTIAL.
FT DISULFID 256 301 POTENTIAL.
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 915 915 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 962 962 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1304 1304 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 2029 AA: 229027 MW: 536A0C794D3DC800 CRC64:

Query Match 13.0%; Score 295.5; DB 1; Length 2029;

```

Query Match          12.8%; Score 292.5; DB 1; Length 837;
Best Local Similarity 21.7%; Pred. No. 1.2e-10;
Matches 101; Conservative 74; Mismatches 139; Indels 151; Gaps 17;

Qy      5  QGRTVTFFPCETKGNPQPAVFWQKEGSQNLFFNPQ---QQPNRCSVSPGDLTITNIQR 61
      :|  :|  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:  |:
Db      224 REEMTLTCKASGSPDPTISWFRNG--KLIBENEKYILKGSNT-----ELTVRNII 273

Qy      62 SDAGYYICQALTVAGSILAKAQLVTDVLTRPPPIILQGPANQTLAVDGTALLKCKATG 121
      :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|

```

```

RESULT 6
NEO1_RAT
ID NEO1_RAT STANDARD; PRT; 1377 AA.
AC P97603;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR (FRAGMENT).
GN NEO1 OR NGN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=97015074; PubMed=8861902;
RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.-Y.,
RA Culotti J.G., Tessier-Lavigne M.;
RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
RL Cell 87:175-185(1996).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U68726; AAB41100.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal.
FT NON TER 1 1

```

FT	SIGNAL	<1	2	POTENTIAL.
FT	CHAIN	3	1377	NEOGENIN.
FT	DOMAIN	3	1074	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1075	1095	POTENTIAL.
FT	DOMAIN	1096	1377	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	36	105	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	135	197	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	232	296	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	324	386	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	405	502	FIBRONECTIN TYPE-III.
FT	DOMAIN	505	598	FIBRONECTIN TYPE-III.
FT	DOMAIN	599	698	FIBRONECTIN TYPE-III.
FT	DOMAIN	704	798	FIBRONECTIN TYPE-III.
FT	DOMAIN	819	919	FIBRONECTIN TYPE-III.
FT	DOMAIN	920	1021	FIBRONECTIN TYPE-III.
FT	DOMAIN	1087	1090	POLY-VAL.
FT	CARBOHYD	42	42	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	179	179	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	295	295	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	439	439	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	458	458	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	608	608	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	684	684	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	878	878	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1377:AA:	150637 MW: E514ED8ADD1A63A9	CRC64:

Query Match 12.5%; Score 284.5; DB 1; Length 1377;
Best Local Similarity 26.3%; Pred. No. 6.5e-10;
Matches 91; Conservative 55; Mismatches 153; Indels 47; Gaps 11;

```

Qy      6  GRTVTFPCETKGNPQPAVFWQKEGSSQNLLFPNQPPQNSRCSVSPGTDLTITNIQRSDAG  65
      |:: || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    232  GQSAVLPCVASGLPAPVIRWMK---NEDVL----DTESSGRLALLAGGSLEISDVETDDAG  285

Qy     66  YYICQALTVAGSILAKAQLEVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDP  125
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    286  TYFCVADNGNKITIEAQELTV-----QVPEFLKQPANIYARESMDVFECEVTGK  339

Qy    126  WISWLKEGFTFPGRDPRATIEQGTILQIKNLRISDTGTYTCTVATSSSGEASWSAVLDVTE  185
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    340  TVKWVKNGDVVTPSDYFKIVKEH-NLQVLGLVKSDEGFYQCI AENDVGNAGAQAQLIILE  398

Qy    186  SGATISKNYDLSDLPGPPSKPQVTDVTKNSVTLWSQPGTPTGLPASAYIIIEAFSQSVSNS  245
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    399  HAPATT-----GPLSPAPRDVVASLVSTRFIKLTWR--TPASDPHG----DNLTYSVF  447

Qy    246  WQTVA-NHVKTT-----LYTVRGLRPNTIYLFMVRAINPKVSVTKXPKQKNGSTW  298
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    448  KEGVARERVENTSQPGEMQVTIQNLMPATVYIFKVMQKNHKGSGESSAPLRVETQPEVQL  507

Qy    299  PLPPPPVQ-----PLPGT-ELEHYAVEQGENGYDSD  328
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    508  PGPAPNIRAYATSPSTIITWTETPLSGNEIGONYKLYMEKGTGDK  553

```

```

RESULT      7
NEO1_CHICK
ID  NEO1_CHICK      STANDARD;      PRT;  1443 AA.
AC  Q90610;
DT  01-OCT-2000 (Rel. 40, Created)
DT  01-OCT-2000 (Rel. 40, Last sequence update)
DT  01-OCT-2000 (Rel. 40, Last annotation update)
DE  NEOGENIN (FRAGMENT).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;
RX  MEDLINE=95105243; PubMed=7806578;
RA  Vielmetter J., Roman J.M., Drever W.J.:

```

```

Qy      6  GRTVTFPCETKGNQPNQPAVFQKEQSGNLLFPNQPPQNSRCSVSPGDLTTITNIQRSDAG  65
      | : ||| | | | | | | | | : : : : | | | : | |
Db    229  GQNAPVFPCVAGFPPTPYVRWTKNGEELI-----TEDSERFALRAGSGLLISDVTEEDVG  282

Qy     66  YYICQALIVAGSILAKAQLETVTLDRPPPIILQGPNQTLAVDGTALLKCKATGDPPLP  125
      | | | : | | : | | : : | | | | : : | | | |
Db    283  TYTCIADNENETIEAQELAV-----QVPEFLKRPANIYAHESMDIVFECEVTKGKPT  336

Qy    126  VISWLKEGFTFPGPRDPRATIOEGTLOIKNLRISDITGYTCVATSSSGEASWSAVLDVTE  185

```

```

Db 337 TVKWKVNGDVVPISDYFKIVKEH-NLQVLGLVKSDEGFQYCAIENDVNGAQAQLIILD 395
Qy 186 -----SGATISKNYDLS-----DLPGPPSKPQVTDVTKNSVTLSSWQPGTPTGLPA 230
Db 396 LDVAIPTLPPTSLTSATNDHLAPATTGPLTAPRDVVATLVSTRFIRLTWR--TPVSDPQ 453
Qy 231 S---AYIEAFSQSVSNWQTIVANHVKITLYTVRGLRPNTIYLFMVRAINPKVSVTXKP 287
Db 454 GDNLTYSIFYIKEGINRREVENTSRPGETQVMIQNLMPETVIVFRVQAQKNKHGHESSAP 513
Qy 288 QKNNGSTWANVPLPPPPVQ-----PLPCT-ELEHYAVEQENGYDSDSWC 331
Db 514 LKVAITPQEVQLPGPAPNIRAYAGSPTSVTVTWETPLSCNGEIQNYKLYMEKGQDSEQ-- 571
Qy 332 PPLPVQ--TYLHQGLEDELEDDRVPTPPVRGVA---SSPAISFGQOSTATLTSPREE 386
Db 572 -DVDVAGLSYITITGLKKYTEYS-----FRVYAYNKHGPGVSTQDVVVRTLSDPV--- 619
Qy 387 MQMPLQASPXFTSSQRPRPTSPFSTDNNTSAALESQSRQRPRPTKKHG 433
Db 620 ----SAAP-----ONLTLEARNKSTMLHWOPPPAGTHSG 650

```

```

RESULT 8
NEO1_MOUSE
ID NEO1_MOUSE STANDARD; PRT; 1493 AA.
AC P97798;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEO1 OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=BRAIN;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is
expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED UBIQUITOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE-NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.
CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
CC TUMOR SUPPRESSOR PROTEIN DCC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstati
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer
CC entities requires a license agreement (See http://www.isb-sib.ch/annou
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; Y09535; CAA70727.1; -.
DR HSSP; P02751; ITTG.
DR MGD; MGI:1097159; NEOL.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 6.
DR PFAM; PF00047; ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
KW Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 1493 NEOGENIN.
FT DOMAIN 37 1136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1137 1157 POTENTIAL.
FT DOMAIN 1158 1493 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 78 147 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 177 239 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 274 338 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 366 428 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 467 564 FIBRONECTIN TYPE-III.
FT DOMAIN 567 660 FIBRONECTIN TYPE-III.
FT DOMAIN 661 760 FIBRONECTIN TYPE-III.
FT DOMAIN 766 860 FIBRONECTIN TYPE-III.
FT DOMAIN 881 981 FIBRONECTIN TYPE-III.
FT DOMAIN 982 1083 FIBRONECTIN TYPE-III.
FT DOMAIN 1149 1153 POLY-VAL.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 MISSING (IN ISOFORM 2).
FT VARSPLIC 863 878 MISSING (IN ISOFORM 3).
FT VARSPLIC 1086 1096 MISSING (IN ISOFORM 4).
FT VARSPLIC 1279 1331 MISSING (IN ISOFORM 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 12.2%; Score 279; DB 1; Length 1493;
Best Local Similarity 24.3%; Pred. No. 1.5e-09;
Matches 112; Conservative 66; Mismatches 191; Indels 92; Gaps 18;

Qy 6 GRTVTFPCETKGNPQPAVFWQKEGSQLNFPNQPPNSRCSVSPGDLTITNIQRSDAG 65
Db 274 GQSAVLPVCGVSGLPAPVVRWMK--NEEVL-----DTESGRVLVLAGGCLIEISDVTEDDAG 327
Qy 66 YYICQALTAVAGSILAKAQLVTDVLTDRPPPIIQGPANQTLAVDGTALLCKATGDPLP 125
Db 328 TYFC--IADNGNKTYEAQAE----LVQVPPGLFKPANITAHESMDIVPECEVTGKPTP 381
Qy 126 VISWLKEGFTFPGDRPRATIQEQTLQIKNLRISDTGTYTCVATSSSGEASWSAVLDVTE 185
Db 382 TVKWKVNGDVVIPSDFNKIVKEH-NLQVLGLVKSDEGFYQCIENDVGNAQAGQLIILE 440
Qy 186 SGATISKNYDLSDLPGPSKQVTDVTKNSVTLSPGPTGTLPASAYIIIEA---FSQSV 242
Db 441 -----HDVAIPTLPPT--SLTSATTDHLA----PATTGPLPSAPRDVVASLVSTRFI 486
Qy 243 SNSQTVAN--HAKTTL-----TVRGLRPNTIYLFMVRA 275
Db 487 KLTWRTPASDPHGDNLTYSVFYTKGCDRREVENTSQGEMQVTIQLNPATVIYFKVMA 546
Qy 276 INPKVSVTQXKPKNGSTWANVLPPLPPVQ-----PLPGT-ELEHYAVE 319
Db 547 QNKHGSGESSAPLRVETQPEVQLPGAPNIRAYATSPTSITVTWETPLSGNGEIQNKLY 606
Qy 320 QQENGYSDSWCPPLPVQTYLHQGLEDELEDDRVPTPPVRGVA---SSPAISFGQST 376
Db 607 YMEKGTDEQDI-DVSSHSTYINGLAKTYEYS-----FRVVAYNKHGPGVSTQDVAV 657

Qy 377 ATLTPSPREEMQPM---LQASPXFTSSQRRPRTSPFSTDSN 414
Db 658 RTLSDVPSSAAPQNLSEVRNSKSIHVHQP----PSSTTQN 694

RESULT 9

CAML_HUMAN

ID CAML_HUMAN STANDARD; PRT; 1257 AA.
AC P32004;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1 OR MIC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92031698; PubMed-1932117;
RA Kobayashi M., Miura M., Asou H., Uyemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of
RT different origin.";
RL Biochim. Biophys. Acta 1090:238-240(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Rosenthal A., Coutelle O., Drescher B.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-92329299; PubMed-1627459;
RA Reid R.A., Hemperly J.J.;
RT "Variants of human L1 cell adhesion molecule arise through alternate
RT splicing of RNA.";
RL J. Mol. Neurosci. 3:127-135(1992).
RN [4]
RP SEQUENCE OF 353-1176 FROM N.A.
RX MEDLINE-92020233; PubMed-1923824;
RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
RT "PCR walking from microdissection clone M54 identifies three exons
RT from the human gene for the neural cell adhesion molecule L1
RT (CAM-L1).";
RL Nucleic Acids Res. 19:5395-5401(1991).
RN [5]
RP SEQUENCE OF 332-371 FROM N.A.
RX MEDLINE-90353957; PubMed-2387585;
RA Djabali M., Mattei M.-G., Nguyen C., Roux D., Demengeot J.,
RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
RT "The gene encoding L1, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [6]
RP SEQUENCE OF 1030-1257 FROM N.A.
RX MEDLINE-91132183; PubMed-1993895;
RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,
RA Stallcup W.B.;
RT "Isolation and sequence of partial cDNA clones of human L1: homology
RT of human and rodent L1 in the cytoplasmic region.";
RL J. Neurochem. 56:797-804(1991).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE-88298876; PubMed-3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule L1.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP VARIANT HSAS TYR-264.
RX MEDLINE-94004956; PubMed-8401576;
RA Jouet M., Rosenthal A., Macfarlane J., Kenwrick S., Donnai D.;

RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus
RT (HSAS).";
RL Nat. Genet. 4:331-331(1993).
RN [9]
RP VARIANT HSAS/MASA LEU-1194.
RX MEDLINE-95187172; PubMed-7881431;
RA Fransen E., Schrandt-Stumpel C., Vits L., Coucke P., van Camp G.,
RA Willems P.J.;
RT "X-linked hydrocephalus and MASA syndrome present in one family are
RT due to a single missense mutation in exon 28 of the L1CAM gene.";
RL Hum. Mol. Genet. 3:2255-2256(1994).
RN [10]
RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.
RX MEDLINE-95004608; PubMed-7920659;
RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,
RA Paterson J., Metzenberg A., Ionasescu V., Temple K., Kenwick S.;
RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked
RT hydrocephalus result from mutations in the L1 gene.";
RL Nat. Genet. 7:402-407(1994).
RN [11]
RP VARIANTS MASA GLN-210 AND ASN-598.
RX MEDLINE-95004609; PubMed-7920660;
RA Vits L., van Camp G., Coucke P., Fransen E., de Boule K.,
RA Reyniers E., Korn B., Poustka A., Wilson G., Schrandt-Stumpel C.,
RA Winter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT L1CAM.";
RL Nat. Genet. 7:408-413(1994).
RN [12]
RP VARIANTS HSAS/MASA S-9; S-121; K-309; F-768; L-941 AND C-1070.
RX MEDLINE-95282776; PubMed-7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Wadelius C., Kenwick S.;
RT "New domains of neural cell-adhesion molecule L1 implicated in
RT X-linked hydrocephalus and MASA syndrome.";
RL Am. J. Hum. Genet. 56:1304-1314(1995).
RN [13]
RP VARIANTS HSAS/MASA Q-184; Q-210; Y-264; R-452; N-598 AND L-1194.
RX MEDLINE-96153146; PubMed-8556302;
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, L1.";
RL Eur. J. Hum. Genet. 3:273-284(1995).
RN [14]
RP ERRATUM.
RA Fransen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
RN [15]
RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE-96057511; PubMed-7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
RA Cassiman J.-J.;
RT "Mutations in L1-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS.";
RL J. Med. Genet. 32:549-552(1995).
RN [16]
RP VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE-97083370; PubMed-8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus.";
RL J. Med. Genet. 33:103-106(1996).
RN [17]
RP VARIANTS HSAS Q-184; V-439--T-443 DEL; C-784 AND L-936--L-948 DEL.
RX MEDLINE-97338664; PubMed-9195224;
RA Macfarlane J.R., Du J.-S., Pepys M.E., Ramsden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus.";
RL Hum. Mutat. 9:512-518(1997).

RN [18]
RP VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE-98180721; PubMed-9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene.";
RL Hum. Mutat. 11:222-230(1998).
RN [19]
RP VARIANT CRASH PRO-632.
RX MEDLINE-98112489; PubMed-9452110;
RA Vits L., Chitayat D., van Camp G., Holden J.J.A., Fransen E.,
RA Willems P.J.;
RT "Evidence for somatic and germline mosaicism in CRASH syndrome.";
RL Hum. Mutat. Suppl. 1:S284-S287(1998).
RN [20]
RP VARIANTS HSAS/MASA THR-219; ARG-335; CYS-386; CYS-473 AND LEU-1224.
RX MEDLINE-98415726; PubMed-9744477;
RA Saugier-Verber P., Martin C., le Meur N., Lyonnet S., Munnich A.,
RA David A., Henocq A., Heron D., Jonveaux P., Odent S., Manouvrier S.,
RA Moncla A., Morichon N., Philip N., Satge D., Tosi M., Frebourg T.;
RT "Identification of novel L1CAM mutations using fluorescence-assisted
RT mismatch analysis.";
RL Hum. Mutat. 12:259-266(1998).
CC -1- FUNCTION: CELL ADHESION MOLECULE WITH AN IMPORTANT ROLE IN THE
CC DEVELOPMENT OF THE NERVOUS SYSTEM. INVOLVED IN NEURON-NEURON
CC ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC. BINDS
CC TO AXONIN ON NEURONS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS IN THE CYTOPLASMIC REGION ARE
CC PRODUCED BY DIFFERENTIAL SPLICING.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF THREE X-LINKED
CC SYNDROMES. 1: HYDROCEPHALUS OWING TO STENOSIS OF THE AQUEDUCT OF
CC SYLVIVUS (HSAS) CHARACTERIZED BY MENTAL RETARDATION AND ENLARGED
CC BRAIN VENTRICLES. 2: MASA SYNDROME WHICH IS CHARACTERIZED BY
CC MENTAL RETARDATION, APHASIA, SHUFFLING GAIT, AND ADDUCTED THUMBS.
CC HAS AN OVERLAPPING PROFILE OF CLINICAL SIGNS WITH HSAS, BUT WITH A
CC MILDER PRESENTATION AND A LONGER LIFE EXPECTANCY. 3: SPASTIC
CC PARAPLEGIA TYPE 1 (SPG1). COLLECTIVELY THESE SYNDROMES ARE ALSO
CC KNOWN AS CRASH SYNDROME, AN ACRONYM WHICH STANDS FOR CORPUS
CC CALLOSUM HYPOPLASIA, PSYCHOMOTOR RETARDATION, ADDUCTED THUMBS,
CC SPASTIC PARAPARESIS, AND HYDROCEPHALUS.
CC -1- DISEASE: DEFECTS IN L1CAM ARE THE CAUSE OF HIRSCHPRUNG DISEASE
CC (HSCR).
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- DATABASE: NAME=L1CAM; NOTE=L1CAM mutation Web Page;
CC WWW="http://hgins.uia.ac.be/dnalab/l1".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X59847; CAA42508.1; -.
DR EMBL; Z29373; CAA82564.1; -.
DR EMBL; M74387; AAA59476.1; -.
DR EMBL; X58775; CAA41576.1; -.

Query Match 11.8%; Score 268; DB 1; Length 1257;
Best Local Similarity 23.8%; Pred. No. 5.7e-09;
Matches 97; Conservative 46; Mismatches 143; Indels 122; Gaps 12;

QY 66 GRIVTFPCETAGNPQPAVFWQEGSNLFFNPQPPQNSRCSVSPGDLTITNIQRSDAG 65
Db 347 GETARLDCQVGRPQPEVTVWRING-----IPVEELAKDQYRIQ-RGALILSNVQSDTM 400
QY 66 YYICQALTVAGSILAKAQLEVTDLTRPPPIILQGPANQTLAVDG-TALLCKATGDPL 124
Db 401 VTQCEARNRHGLLANAYIYVQL-----PAKILLADNTYMAVGSGSTAYLLCKAFAPV 455

```
Qy      6 GRTVFPFCETKGNPQPAVFQWQ-----EGSONLLFPNQPPQNRSRVSPTGLTIITN 58  
       |:| || | | : | | | | : | : | : | | | :  
Db    263 GDQVVLPCVASGLPTPTIKWMNEEALDESSERLV-----LLAGGSLEISD 309  
  
Qy     59 IQRSDAGYYICQALTAVGSLAKAQLEVTDLTRDPPIILQGPNQTLAVDGTALLCKC 118  
       : || | | | : | | : | | : | | : | : | :  
Db    310 VTDEDAGTYFCIADNGNETIEAQELTV-----QAQEFLKQPNTNIYAHESMDIVFECE 363  
  
Qy    119 ATGDPLPVISWLKEGFTFPGDRPRATIQEQSTLQIKNLRI SDTGYTCVATSSSGEASWS 178  
       || | : | : | | | : | : | | | | : | : | : | : |  
Db    364 VTGKPTPTVKPVKNQDMVIPSDYPFKIVEH-NLQVLGLVKSDEGFQCIAENDVGNAQAG 422  
  
Qy    179 AVLDTVESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLWSQPGTPGTLPAS---AYII 235  
       | | : | : | | | : | : | : | : | : | : | : | :  
Db    423 AQLIILEHAPATT----GPLSPAPRDVVASLVSTRFIKLWR--TPASDPHGDNLIYSV 475  
  
Qy    236 EAFSQSVSNWSQTVAHVKTTLYTVRGLRPNTIYLFMVRAINPKVSVTQXKPQKNNGSTW 295  
       : : : : | : | : | : | : | : | : | : | : | :  
Db    476 FYTEKGIAREREVENTSHPGEMQVTIQLNPMPATVIYFRVMAQKHGSGESSAPLRVETQPE 535  
  
Qy    296 ANVPLPPPPVQ-----PLPGT-ELEHYAVEQQENGYS D 328  
       : | | | : | : | : | : | : | : | : | :  
Db    536 VLPGCAPNLRKAYASPTSITVTWEIPVSGNGEIIONKLYIMEKGTDKE 584
```

FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	587	587	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	725	725	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	776	776	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	824	824	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	848	848	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	875	875	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	968	968	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	978	978	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1022	1022	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1030	1030	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1107	1107	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1180	1183	MISSING (IN SHORT ISOFORM)
FT				(BY SIMILARITY)
SQ	SEQUENCE	1260 AA;	140968 MW;	22B5E7001CB2A538 CRC64;

Query Match 11.6%; Score 265.5; DB 1; Length 1260;
Best Local Similarity 21.3%; Pred. No. 8.1e-09;
Matches 122; Conservative 70; Mismatches 197; Indels 183; Gaps 19;

```

Qy      6  GRTVFPCEKTKGNPQAVFWQKEQSSQLFFNPQQPQSRCSVSPGTGDLTITNQRSADG  65
      | |   | : | | | : | : | : | : | |   | : | : |
Db 346  GETARLDCQVQGRPQPEITWRING-MSMETVNKDQKYRIE-----QGSILNSNQPTDM  399

Qy      66  YYICQALTVAGSILAKAQLEVTDVLTDRPPPIILQGPAQOT-LAVDG-TALLKCKATGDP  123
      | : | | : | | : | : | | : | | | | | | |
Db 400  VTQCEARNQBGHLLANAYITVVQL----PARILTKD--NQYTMAVEGSTAYLLCKAFGAP  453

Qy      124  LPVISWLKEGFTPPGRDPRATIQEGQTLQIKNLRISDTGTTCVATSSSGEASWSAVLDV  183
      : | : | | | : | | | | | : | : | | | | | |
Db 454  VPSVQWLDEEGTTLQDERFFPYANGTSLIRDLQANDTGRYFCCAANDQNNVTILANLQV  513

Qy      184  TES-----G  187
      | :
Db 514  KEATQITQGPRAIEKKGARVTTQCAQSTFDSLQASITWRGDRGLQERGSDKYFIEDG  573

Qy      188  ATISKNYDLSD-----LPGPPSKPQVTD--VTKNSVTLS  219
      : : | | | | | | | | | : : | |
Db 574  KLVIQSLDYSQGNYSVCVASTELDEVESRAQLLVGSPGVPVHLELSDRHLLKQSQVHLS  633

Qy      220  WQPGTPTGTLPASAYIEAFSQSVS-NSW----QTVANHVKTTLITVRGLRPNTIYLFMVR  274
      | | | | | | | | | | | | | | | |
Db 634  WSPAEDHNSPIEKYDIEFEDKEMAPEKWFSLGKVQNGQSTTL----KLSPYVHYTFRVT  689

Qy      275  AIN-----PKVSVTQXKPQKN-----NGSTWANVPLPPPPVQPLPGTELE-HY  316
      | | | : | : | : | : | : | : | : |
Db 690  AINKYGPGEPSPVSESVTPEAAEKNPVDVRGEGNETNMVITWKPLRWMWNAPIQIY  749

Qy      317  AVEQQENGYD-----SDSWCPPLPVQTYLHQGLEDEEDDDRVPVTPVRGVASSPA  368
      | | | : | : | : | : | : | : | : |
Db 750  RVQWRPQKGQETWRKQTVSDPFLVMSNTSTFVPYBIKVQAVNNQKGKPEP-----  799

Qy      369  ISFGQOSTATLTPSPREEMQPLQASFXFTSSQ---RPRPT-----  406
      | | | : | : | | | | | |
Db 800  ----QVTIGYSGEDYIPQVSPEDITIFNSSITVLVRWVPVLAQVKGHLKGYNVTYWK  854

Qy      407  -----SPFSTDSNTSAAALSQSRP  425
      : : | : | |
Db 855  GSORKSKSRHIBKSHIVPANTTSAILSGLRP  886

```

```

RESULT 12
CAML_RAT
ID CAML_RAT STANDARD; PRT; 1259 AA.
AC Q05695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE L1 PRECURSOR (N-CAM L1).
GN L1CAM OR CAML1

```


Query Match 11.4%; Score 259; DB 1; Length 1447;

```

RESULT 14
NRG_DROME
ID NRG_DROME STANDARD; PRT; 1239 AA.
AC P20241; Q24414;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGLIAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RX SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RP MEDLINE=90030418; PubMed=2805067;
RA Bieher A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglian: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RL LI."
RL Cell 59:447-460(1989).
RN [2]
RX SEQUENCE OF 1182-1239 FROM N.A.
RP MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bieher A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglian."
RL Neuron 4:697-709(1990).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RP MEDLINE=94213741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bieher A.J., Bjorkman P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglian at 2.0 A."
RL Neuron 12:717-731(1994).
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL

```


DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 4.
DR PFAM; PF00047; ig; 6.
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1012 AXONIN-1.
FT PROPEP 1013 1040 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 54 118 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 148 216 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 254 313 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 341 402 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 433 495 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 523 594 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 606 612 GLY/PRO-RICH.
FT DOMAIN 611 706 FIBRONECTIN TYPE-III.
FT DOMAIN 714 809 FIBRONECTIN TYPE-III.
FT DOMAIN 816 908 FIBRONECTIN TYPE-III.
FT DOMAIN 917 1003 FIBRONECTIN TYPE-III.
FT SITE 794 796 CELL ATTACHMENT SITE (BY SIMILARITY).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 1012 1012 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1040 AA; 113393 MW; 254E78DD3C28EFB6 CRC64;

Job time: 1292 sec

Query Match 11.2%; Score 256; DB 1; Length 1040;
Best Local Similarity 24.4%; Pred. No. 2.4e-08;
Matches 108; Conservative 59; Mismatches 151; Indels 124; Gaps 22;

Qy 4 AQGRVTTPCETKGNPQPAVFWQKEGSSQNLFPNQPPQNSRCSVSPTGDLITITNIQRSD 63
|:| : ||: : | : | | | : : : ||:| : | | | | | |
Db 431 ARGGEILIPQCPRAAPKAVVLWSK-GTEILV-----NSSRVTVPDGLTIIRNISRS 482

Qy 64 AGYYICQALTIVAGSILAKAQLEVTDLVTRPPPIILQGPANQTLAVDGTALLKCKATGDP 123
| | | | : : | | : | : | : | : : : | : | : |
Db 483 EGKYTC---FAENFMGKA--NSTGILSVRDATKITLAPSSADINLDNLTQCHASHDP 536

124 LP--VISWLKEGFTF---PGRDPRAT-IQEQ-GTLQIKNLRIISDTGTYTCVATSSSGEA 175
: | : | : | | | : | : | : | : | : | : | : | : |
Db 537 TMDLITFTWTLDDFPIDFDPKGGHYRRTNVKETIGDLTILNAQLRHGGKYTCMA----- 589

Qy 176 SWSAVLDVTESGATISKNYDLSLPGPPSPKPQVTDVTKNSVTLQWPGTPTGLPASAYII 235
|:| | | : | : | : | | | : : | : | : | : | : |
Db 590 --QTVVDSASKEATVL---VRGPPGPPGVVVRDIDGTTIQLSWSRGFDNHSPIAKYTL 643

Qy 236 EAFSQSVSNWQTV---AN---HVKTILYTVRGLRPNTIYLFMVRIN----- 277
: | : : | : | | : : | | | | | | | |
Db 644 QARTPP-AGKWKQVRTNPANIEGNAETA--QVLGLTPWMDYEFRIASNILGTGEPSPGPS 700

Qy 278 PKVSVTQXKPKQ-----KNGGST-WAN 297
| : : | : : : | : | : | : | : | : | : |
Db 701 SKINTREAPSVAPSGLSGGGAPGELIVNWTNPSREYQNGDGFYLLSPRRQGSTHWQT 760

Qy 298 VPLPPPPVQPLPGTELEHYAVEQQENGYSDSWCPLP----VQTYLHQG-----LE 345
: | : : : | : | : | : | : | : | : | : | : |
Db 761 A-----RVPGADAQYFV-----YSNESVRPVTPEVKIRSYNRRGDGPESLTALV 805

Qy 346 DELEEDDRVPTTP-VRGVASS 366
| : | : | : | : | : | : | : | : | : |
Db 806 YSAEEPRVAPTKVWAKGVSSS 827

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:54:03 ; Search time 559.88 Seconds
(without alignments)
90.856 Million cell updates/sec

Title: US-09-540-245A-19
Perfect score: 2280
Sequence: 1 QIVAQGRITVTPCETKGNPQ.....TSAALSQSRPRPTKKHKG 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp Vertebrate:
14: sp_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1479	64.9	285	4	Q43608	Q43608 homo sapien
2	1409	61.8	1060	11	Q9QZ13	Q9QZ13 rattus norv
3	913	40.0	1651	4	Q9Y6N7	Q9Y6N7 homo sapien
4	911	40.0	1612	11	O89026	O89026 mus musculu
5	911	40.0	1651	11	O55005	O55005 rattus norv
6	755.5	33.1	1344	11	Q9Z214	Q9Z214 mus musculu
7	579	25.4	1273	5	O44928	O44928 caenorhabdi
8	545	23.9	1395	5	O44924	O44924 drosophila
9	545	23.9	1395	5	Q9W213	Q9W213 drosophila
10	478	21.0	823	5	Q9VQ10	Q9VQ10 drosophila
11	417	18.3	859	5	Q9VPZ6	Q9VPZ6 drosophila
12	388	17.0	874	5	O01632	O01632 caenorhabdi
13	336.5	14.8	2221	5	Q9U1M1	Q9U1M1 drosophila
14	331.5	14.5	2222	5	O97394	O97394 drosophila
15	330	14.5	1788	13	Q9IAJ0	Q9IAJ0 xenopus lae
16	330	14.5	1904	11	Q64699	Q64699 mus musculu
17	325	14.3	1499	13	Q90815	Q90815 gallus gall
18	323.5	14.2	1501	11	Q9QW00	Q9QW00 rattus sp.
19	323.5	14.2	1863	11	Q64605	Q64605 rattus norv

20	322.5	14.1	1948	4	Q13332	Q13332 homo sapien
21	319.5	14.0	475	4	O75255	O75255 homo sapien
22	316	13.9	1894	11	Q64487	Q64487 mus musculu
23	314	13.8	1502	4	Q9UM81	Q9UM81 homo sapien
24	308.5	13.5	1896	13	Q9IAJ1	Q9IAJ1 xenopus lae
25	305.5	13.4	1898	11	Q64604	Q64604 r protein-t
26	300.5	13.2	1598	4	Q9P214	Q9P214 homo sapien
27	297.5	13.0	1232	13	Q90284	Q90284 carassius a
28	297	13.0	1264	5	P91767	P91767 manduca sex
29	297	13.0	1887	11	Q9QW67	Q9QW67 rattus sp.
30	296.5	13.0	1026	11	Q62845	Q62845 rattus norv
31	296.5	13.0	1277	13	Q98902	Q98902 fuqu rubrip
32	288	12.6	2016	5	Q9V4J9	Q9V4J9 drosophila
33	288	12.6	2016	5	Q9NBAL	Q9NBAL drosophila
34	287.5	12.6	1180	4	O15051	O15051 homo sapien
35	285.5	12.5	1299	4	Q92823	Q92823 homo sapien
36	285.5	12.5	1496	4	Q92626	Q92626 homo sapien
37	284.5	12.5	1377	11	P97603	P97603 rattus norv
38	283.5	12.4	1099	11	P97527	P97527 rattus norv
39	282	12.4	1443	13	Q90610	Q90610 gallus gall
40	281.5	12.3	1236	4	Q9UHI3	Q9UHI3 homo sapien
41	281.5	12.3	1308	4	Q9UHI4	Q9UHI4 homo sapien
42	281	12.3	793	11	O70246	O70246 mus musculu
43	279	12.2	1493	11	P97798	P97798 mus musculu
44	278	12.2	920	4	Q9P232	Q9P232 homo sapien
45	276	12.1	1028	11	Q62682	Q62682 rattus norv

ALIGNMENTS

RESULT 1
O43608
ID O43608 PRELIMINARY; PRT; 285 AA.
AC O43608;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ROUNDABOUT 2 (FRAGMENT).
GN ROBO2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040991; AAC39576.1; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 1.
DR PFAM; PF00047; lg; 2.
FT NON_TER 1 1
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 30606 MW; 05DF916A3DBA96C6 CRC64;

Query Match 64.9%; Score 1479; DB 4; Length 285;
Best Local Similarity 100.0%; Pred. No. 8.9e-104;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIVAQGRITVTPCETKGNPQPAVFVWQKEGSGNLLFPNQPPQPNRCSVSPTGLTITNIQ 60
|||||
Db 1 QIVAQGRITVTPCETKGNPQPAVFVWQKEGSGNLLFPNQPPQPNRCSVSPTGLTITNIQ 60

Qy 61 RSDAGYYICQALTAVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
|||||
Db 61 RSDAGYYICQALTAVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120

Qy 121 GDPLPVISWLKEGFTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASNSAV 180
|||||
Db 121 GDPLPVISWLKEGFTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASNSAV 180
Qy 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTLPASAYIEAFSQ 240
|||||
Db 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTLPASAYIEAFSQ 240
Qy 241 SVSNSWQTVANHVKTLLYTVRGLRPNTIYLFMVRAINPKVSVTQ 284
|||||
Db 241 SVSNSWQTVANHVKTLLYTVRGLRPNTIYLFMVRAINPKVSVTQ 284

RESULT 2

Q9QZI3

ID Q9QZI3 PRELIMINARY; PRT; 1060 AA.
AC Q9QZI3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
REPULSIVE GUIDANCE RECEPTOR (FRAGMENT).
Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99200391; PubMed=10102268;
RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
RA Tessier-Lavigne M., Kidd T.;
RT "Slit proteins bind Robo receptors and have an evolutionarily
RT conserved role in repulsive axon guidance.";
RL Cell 96:795-806(1999).
DR EMBL; AF182037; AAF04558.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Receptor.
FT NON_TER 1060 1060
SQ SEQUENCE 1060 AA; 116790 MW; C4BC8C11E8542DA4 CRC64;

Query Match 61.8%; Score 1409; DB 11; Length 1060;
Best Local Similarity 64.7%; Pred. No. 8.1e-98;
Matches 303; Conservative 28; Mismatches 81; Indels 56; Gaps 10;

Qy 1 QIVAQGRVTTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQPNRCSVSPTGDLTITNIQ 60
|||||
Db 326 QIVAQGRVTTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQPNRCSVSPTGDLTITNIQ 385
Qy 61 RSDAGYYICQALTIVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
|||||
Db 386 RSDAGYYICQALTIVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 445
Qy 121 GDPLPVISWLKEGFTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASNSAV 180
|||||
Db 446 G-PLPVISWLKEGFTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASNSAV 504
Qy 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTLPASAYIEAFSQ 240
|||||
Db 505 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTLPASAYIEAFSQ 564
Qy 241 SVSNSWQTVANHVKTLLYTVRGLRPNTIYLFMVRAINPKVSVTQXKPKQNNGTWVANPL 300
|||||
Db 565 SVSNSWQTVANHVKTLLYTVRGLRPNTIYLFMVRAINPKVSVTQXKPKQNNGTWVANPL 620
Qy 301 PPP-----PVQPLPGTELEHYAVEQENGYSWCPPLPVQT 338

Db 621 SPPAQGVDRHQVQKELGDTVRLHNPVLTPTTVQVTVTDRO-----PQFIQG 669
Qy 339 Y-----LHQGLEDEL--EEDDDRVPTTPVVRGVASSPAISFGQSTATLTPSPR-EEMQPM 390
|||||
Db 670 YRVMYRQTSGLQASTVWQNLDAKVPT-----RSAVLVNKKGVTEIKVRPYFNEFGM 724
Qy 391 LQASPXFTSSQR-----PRPTSPFSTDSNTSAAALSQSQRPRTKHKH 433
|||||
Db 725 DSESKTIRTTEAPSAPPQSVTVLVGSHNSTSISVSWDPPADHQNG 772

RESULT 3

Q9Y6N7

ID Q9Y6N7 PRELIMINARY; PRT; 1651 AA.
AC Q9Y6N7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROUNDABOUT 1.
GN ROBO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF040990; AAC39575.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1651 AA; 180928 MW; 9D98CD7CAB73074D CRC64;

Query Match 40.0%; Score 913; DB 4; Length 1651;
Best Local Similarity 23.2%; Pred. No. 2.8e-60;
Matches 258; Conservative 54; Mismatches 118; Indels 684; Gaps 12;

Qy 1 QIVAQGRVTTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQPNRCSVSPTGDLTITNIQ 60
|||||
Db 360 QIVAQGRVTTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQPNRCSVSPTGDLTITNIQ 419
Qy 61 RSDAGYYICQALTIVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
|||||
Db 420 RSDAGYYICQALTIVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 479
Qy 121 GDPLPVISWLKEGFTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASNSAV 180
|||||
Db 480 GSPVPTILWRKDGVLVSTQDSRIKLENGVLQIRYAKLGDGTGYTCIATSPSGEATWSAY 539
Qy 181 LDVTESGATISKNYDLSLPGPPSKPQVTDVTKNSVTLSWQPGTLPASAYIEAF 238
|||||
Db 540 IEVQEFQVPPVPPRPTDNLIPSAKPEVTDVSRNTVLSWQPNLNSGATPTSYIEAF 599
Qy 239 SVSNSWQTVANHVKTLLYTVRGLRPNTIYLFMVRAIN----- 277
|||||
Db 600 SHAGSGSWQTVANVKTETSAIKGLKPNAYILFLVRAANAYGISDPSQISDPVKTVQVLP 659
Qy 278 ----- 277
Db 660 TSQGVDRKQVQRELGNVHLHNPVLTSSSIEVHWTVQSQYIQYKILYRPSGANHG 719
Qy 278 -----PK 279
Db 720 ESDWLVEFVRTPAKNSVPIPLRKGVNYEIKARPFNEQGDSEIKFAKTELEAPSAPP 779

```

RESULT      4
O89026
ID   O89026          PRELIMINARY;          PRT;  1612 AA.
AC   O89026;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DUTT1 PROTEIN.
GN   ROBO1 OR DUTT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Wu M.C., Lowe N., Fordham R., Rabbitts P.;
RT   "The mouse homologue of human DUTT1/H-robo1 gene: protein sequence and
RT   chromosomal location.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y17793; CAA76850.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:1274781; Robo1.

```

[illegible]

Db 1216 YSHQSTATLTPSPQEEELQPMQLQDCPEDLGHMPPDRRRQPVSPPPPPRPISPHTYGYI 1275
Qy 396 ----- 395
Db 1276 SGPLVSDMDTDAPEEEEDADMEVAKMQTRRLRLRGLEQTPASSVGDLESSVTGSMINGW 1335
Qy 396 -----XFTS 399
Db 1336 GSASEEDNISSGRSSVSSSDGSFFTDADFAQAVAAAAAYAGLKVARRQMQDAAGRHHFHA 1395
Qy 400 SQRPRPTSPFSTDSNTSAAALSQSQRPRTKKHKG 434
Db 1396 SQCPRTSPVSTDSNMSAVVIQKARPAKQKHQPG 1430

RESULT 5

O55005

ID O55005 PRELIMINARY; PRT; 1651 AA.
AC O55005;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TRANSMEMBRANE RECEPTOR ROBO1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
DR EMBL; AF041082; AAC39960.1; -.
DR HSSP; P56276; 1TLK.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
KW Transmembrane.
SQ SEQUENCE 1651 AA; 180746 MW; FA2452DD46E186B7 CRC64;

Query Match 40.0%; Score 911; DB 11; Length 1651;
Best Local Similarity 23.0%; Pred. No. 4e-60;
Matches 256; Conservative 57; Mismatches 117; Indels 684; Gaps 12;

1 QIVAGRTVTFPCETKGNPQPAVFWQKEGSQLNLLFPNQPPQPNRSRCSVPTGDLTITNIQ 60
Db 360 QVVALGRTVTFQCEATGNPQPAIFWRREGSQLNLLFPNQPPQSSRSVSTGDLTITNIQ 419
Qy 61 RSDAGYICQALTAVGSILAKAQLVDTVLTDRPPPIILQGPANQTLAVDGTALLKCKAT 120
Db 420 RSDVGYICQTLNVAGSIITKAYLEVTDVIADRPVPIRQGPVNVQTVAVDGTLLTLCVAT 479
Qy 121 GDPLPVIWNLKEGFTFPGRDPRATIEQEGTLQIKNLRIISDTGTYTCVATSSSGEASWAV 180
Db 480 GSPVPTILWRKDGVLVSTQDSRIKQLESGLVQIRYAKLGDGTGRYTCTASTSGEATWSAY 539
Qy 181 LDVTEGATIS--KNYDLSDLPGPPSPQVTDVTKNSVTLWQPGTPTLPASAYIIIEAF 238
Db 540 IEVQEGFVQPPRPTDNLIPSAPSKPEVTDVSKNTVLLWQPNLNSGATPTSYIIIEAF 599
Qy 239 SQSVSNWQTVANHVKTTLTYTVRGLRPNTIYLFMVRIN----- 277
Db 600 SHASGSSWQTVANVKTETFAIKGLKPAIYLFVRAANAYGISDPSQISDPVKTQDVPP 659
Qy 278 ----- 277
Db 660 TTQGVHDHQQVRELGNVVLHLNPTILSSSSVEHWTVDQSQYIYQYKILYRPSGASHG 719

Qy 278 -----PK 279
Db 720 ESEWLVEVVRTPTKNSVVIPLDRKGVNYEIKARPFNEFQAGADSEIKFAKLEERPSAPP 779
Qy 280 VSVTQXK----- 286
Db 780 RSVTVSKNDGNGTAILVTWQPPPEDTQNGMVQYKVCWLGNETRYHINKTVDGSTFSVVI 839
Qy 287 -----PO----- 288
Db 840 PFLVPGIRYSVEVAASTGAGPGVKSEPFQIQLDSHGNPVSPEDQVSLAQQISDVVKQPAF 899
Qy 289 -----KNNG----- 292
Db 900 IAGIGACWILMVFSIWLYRHRKRKNGLSSTYAGIRKVPSTFTPTVTYQRGGEAVSSG 959
Qy 293 -----STWAN----- 297
Db 960 GRPGLLNISEPATQPLADTWNTGNSHNDSCINCTASNGNSDNLTTYSRPADCIANY 1019
Qy 298 ----- 297
Db 1020 NNQLDNKQTNLMLPESTVYGVVDLSNKNINEMKTFNSPNLKDGRFVNPSPGQTPYATTQLI 1079
Qy 298 -----VP-- 299
Db 1080 QANLINNMNNGGDSSEKHWKPPGQKQEVAPIQYNIMEQNLKNDYRANDTILPTIPYN 1139
Qy 300 -----LPPPPVQPLPGTEL 313
Db 1140 HSYDQNTGGSYNSSDRSGSSGSGHKKGARTPKAPKQGGMNWADLLPPPPAHPPPHSNS 1199
Qy 314 EHYAVEQENGYSDSWCPPLPVQTYLHQGLEDELEEDD-DRVPTPPVRGVASSP-AISF 371
Db 1200 EYSMSWDES-YDQEMPCPVPPARMYLQ---DELEEEEAERGPTPPVRGAASSPAVSY 1255
Qy 372 GQOSTATLTPSPREEMQPMLOASP----- 395
Db 1256 SHQSTATLTPSPQEEELQPMQLQDCPEDLGHMPPDRRRQPVSPPPPPRPISPHTYGYIS 1315
Qy 396 ----- 395
Db 1316 GPLVSDMDTDAPEEEEDADMEVAKMQTRRLRLRGLEQTPASSVGDLESSVTGSMINGW 1375
Qy 396 -----XFTS 400
Db 1376 SASEEDNISSGRSSVSSSDGSFFTDADFAQAVAAAAAYAGLKVARRQMQDAAGRHHFHA 1435
Qy 401 QRPRPTSPFSTDSNTSAAALSQSQRPRTKKHKG 434
Db 1436 SQCPRTSPVSTDSNMSAAVIQKARPTKKQKHQPG 1469

RESULT 6

Q922I4

ID Q922I4 PRELIMINARY; PRT; 1344 AA.
AC Q922I4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE RIG-1 PROTEIN.
GN RBIG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan S.-S.F., Cox L.A., Dasika G.K., Lee E.Y.-H.P.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060570; AAD11628.1; -.
DR HSSP; P56276; 1TLK.

Db 419 CSAFSVVDSSVVRVFLQVSSV-DERPPPIIQIGPANQTLPGKSVATLPCRATGNPSPRIK 477
Qy 129 WLKEGFTFPGRDPRATIQEQGLTIQIKNLRISDTGTYTCVATSSSGEASWSAVLDVTESGA 188
Db 478 WFHDGHAVQAGN-RYSIIQGSRLRVDLQLSDSGTYTCTASGERGETSWAATLTVEKPGS 536
Qy 189 T-ISKNYDLSLPGPPSKPQVTDVTKNSVTLW----QPQTPGTLPASAYIIIEAFSQSV 242
Db 537 TSLHRAADPSTYPAPPPTPKVLNVSRTSISLRWAKSQEKPGAVG--PIIGTYVEYFSPDL 594
Qy 243 SNSWQTVANHVKTITLYTVRGLRPNTIYLFMVRINPK-VSV 282
Db 595 QTGWIVAHRVGDGTQVTISGLTPGTSYVFLVRAENTQGISV 635

RESULT 9

Q9W213

ID Q9W213 PRELIMINARY; PRT; 1395 AA.
AC Q9W213;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ROBO PROTEIN.
GN ROBO.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler J., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003458; AAF46887.1; -.

DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0005631; robo.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13F0 CRC64;

Query Match 23.9%; Score 545; DB 5; Length 1395;
Best Local Similarity 40.9%; Pred. No. 1.1e-32;
Matches 115; Conservative 47; Mismatches 105; Indels 14; Gaps 7;

Qy 9 VTFPCETKGNPQPAVFWQKEGSQLNLPNPQPPNSRCSVSPGDLTITNIQRSDAGYI 68
Db 362 VQLPCMASGNPPSVFVTKGVSILMFPN---SSHGQHVADGTLQITDVRQEDEGYIV 418
Qy 69 CQALTAVAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDLPLVIS 128
Db 419 CSAFSVVDSSVVRVFLQVSS-LDERPPPIIQIGPANQTLPGKSVATLPCRATGNPSPRIK 477
Qy 129 WLKEGFTFPGRDPRATIQEQGLTIQIKNLRISDTGTYTCVATSSSGEASWSAVLDVTESGA 188
Db 478 WFHDGHAVQAGN-RYSIIQGSRLRVDLQLSDSGTYTCTASGERGETSWAATLTVEKPGS 536
Qy 189 T-ISKNYDLSLPGPPSKPQVTDVTKNSVTLW----QPQTPGTLPASAYIIIEAFSQSV 242
Db 537 TSLHRAADPSTYPAPPPTPKVLNVSRTSISLRWAKSQEKPGAVG--PIIGTYVEYFSPDL 594
Qy 243 SNSWQTVANHVKTITLYTVRGLRPNTIYLFMVRINPK-VSV 282
Db 595 QTGWIVAHRVGDGTQVTISGLTPGTSYVFLVRAENTQGISV 635

RESULT 10

Q9VQ10

ID Q9VQ10 PRELIMINARY; PRT; 823 AA.
AC Q9VQ10;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG5481 PROTEIN.
GN CG5481.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

RESULT 11
Q9VPZ6
ID Q9VPZ6 PRELIMINARY; PRT;
AC Q9VPZ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

```
QY      3 VAQGRVTVFPCETKGNPQPAVFWEQKESQNLFPNQPPQNSRCSVSPGDT---LIITN 58
      | | | : | | | | : | | : : | : | | : | : | |
Db     269 VELGADTSPFCRAIGNPKPTIFWTIKNNSTLIFPGAP--PLDRFHSLNTEEGSHLTLTR 326

QY     59 IQRSADAGYII-CQALTAVGASILAKAEVTDVLTDRPPIIILQGPANQTLAVDGTALLKC 117
      | | | | | : | | : | : | | | | | | | | | | |
Db    327 FORTKDKLVILCNAMNEVASITSRVOLSLS-DSOEDRPPIIISGPNVNTLPKISLATLOC 385
```

Qy	86	VTDVLTDRPPPIILGQPNQTLAVDGTALLKCKATGDP LPVISWLKEGFTFPGRPRATI	145
		: : : : : : : :	
Db	20	VTGNTPAKPPTTIEHGHNQTLWVGSSAILPQCASGXPTPGISWLRDLGLPIDITSRI SQ	79
Qy	146	BEQGTLQIKNRLISDTGYTYCVATSSSGEASWSAVLDVTE--SGATISKNYLDSLPLGPP	203
		: : : : : : : : :	
Db	80	HSTGSLHIADLKPKPTGVYTCIARNDEGESTWSASLVEDHTSNAQFVRMPDSPNFSPS	139
Qy	204	SKPQVTDVTKNSVTSLSWQ-PGTPGTLPASAYIIIEAFSQSVSNSTQTVANHVKTLYTVRG	262
		: : : : : : : : : :	
Db	140	TQPIIVNVITDEVELHWNAPSTSAGGPITGYIIQYYSPLDGQTWFNIPIDVASTEYIRKG	199
Qy	263	LRPNTIYLEMVRAIN-----PKVS--VTQXKP	287
		: : : : : :	
Db	200	LKPSHSYMFVIRAENEKGIGTPSVSSALVTTSKP	233

```

RESULT 13
Q9U1M1
ID Q9U1M1 PRELIMINARY; PRT: 2221 AA.
AC Q9U1M1; Q9UB11; Q9W5D9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SDK PROTEIN (EG:BACR19J1.1 PROTEIN).
GN SDK OR EG:BACR19J1.1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Borkova D., Minana B., Kafatos F.C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RA Benos P.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle-C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G.,

```

Query Match 17.0%; Score 388; DB 5; Length 874;
Best Local Similarity 38.8%; Pred. No. 3.9e-21;
Matches 83; Conservative 38; Mismatches 81; Indels 12; Gaps 4;

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE OF 1704-2221 FROM N.A.
RC STRAIN=BERKELEY;
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
RA Park S., Sequeira A., Sethi H., Sair E., Svirskas R.R., Weinburg T.,
RA Celnikier S.E.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132792; CAB65848.1; -.
DR EMBL; AE003418; AAF45541.1; -.
DR EMBL; AF132195; AAD34783.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0021764; sdk.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 13.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
KW Polymorphism.
FT VARIANT 51 51 Y -> CYAD (IN STRAIN OREGON-R).
SQ SEQUENCE 2221 AA; 245963 MW; 67451AD6A57F06F0 CRC64;

Query Match 14.8%; Score 336.5; DB 5; Length 2221;
Best Local Similarity 28.0%; Pred. No. 9.2e-17;
Matches 109; Conservative 56; Mismatches 141; Indels 83; Gaps 16;

Qy 6 GRVTTFPCETKGNPQPAVFWQKEGSONLLFPNQPPQPNRSCSVSPTGDLTITNIQRSDAG 65
Db 466 GKDATISCRVAGSPNPNITW-----IYNETQLVDISSRVQILESGDLLISNIRSDAG 518
Qy 66 YYICQALT VAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDP-L 124
Db 519 LYICVRANEAGSVKGEAYLSVL-VRTQ-----IIQPPVDTTVLLGLTATLQCKVSSDPV 572
Qy 125 PV-ISWLKEG--FTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASWSAVL 181
Db 573 PINIDWYREGQSSTPISNSQRIGVQADGLEIQA VRASDVGSYACVVTSPGGNETRAARL 632
Qy 182 DVTESGATISKNYDLSDLPGPSKPQV---TDVTKNSVTLSWQPGTGPLPASAYII--- 235
Db 633 SVIE-----LPPPPSNVVERLPEPQQASINVSPTPGFDGNSPISKFIQRR 679
Qy 236 -----EAFSQSVSN---SWQT VANHVKT--TLYTVRGLRPNTIYLFMVRAIN----- 277
Db 680 EVSELEKFGVGPVDPDLLNWITELSNVSADQRWILLENKAAATVYQFRVSAVN---RVGEG 736
Qy 286 KPQKNNGSTWAVNPL-----PP-----PPVQPLPGTELEHYAVEQQE 322
Db 737 SP--SEPSNVVLPQEAAPSGPPVFGVGSARSMSSEIITQWQPLEEHRNGQILGYILRYL 794
Qy 323 NGYDSDSWCPPLPVQTYLHQGLEDELEED 351
Db 795 FGYNVWPWS-----YQNITNEAQRN 814

RESULT 14
097394
ID 097394 PRELIMINARY; PRT; 2222 AA.

AC 097394;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SIDEKICK PROTEIN.
GN SDK.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen D.N.T., Liu Y., Litsky M.L., Reinke R.;
RT "Sidekick, a member of the immunoglobulin superfamily, is required for
RT pattern formation in the *Drosophila* eye."
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88578; AAD09632.1; -.
DR HSSP; P56276; 1TLK.
DR FLYBASE; FBgn0021764; sdk.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 13.
DR PFAM; PF00047; ig; 5.
DR PRINTS; PR00014; FNTYPEIII.
SQ SEQUENCE 2222 AA; 246174 MW; 18853CCAF98D3BC2 CRC64;

Query Match 14.5%; Score 331.5; DB 5; Length 2222;
Best Local Similarity 27.3%; Pred. No. 2.2e-16;
Matches 107; Conservative 58; Mismatches 138; Indels 89; Gaps 15;

Qy 6 GRVTTFPCETKGNPQPAVFWQKEGSONLLFPNQPPQPNRSCSVSPTGDLTITNIQRSDAG 65
Db 469 GKDATISCRVAGSPNPNITW-----IYNETQLVDISSRVQILESGDLLISNIRSDAP 521
Qy 66 YYICQALT VAGSILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKATGDP-L 124
Db 522 LYICVRANEAGSVKAEAYLSVL-VRTQ-----IIQPPVDTTVLLGLTATLQCKVSSDPV 575
Qy 125 PV-ISWLKEG--FTFPGDRPRATIEQEGTLQIKNLRISDTGTYTCVATSSSGEASWSAVL 181
Db 576 PYNIDWYREGQSSTPISNSQRIGVQADGLEIQA VRASDVGSYACVVTSPGGNETRAARL 635
Qy 182 DVTESGATISKNYDLSDLPGPSKPQV---TDVTKNSVTLSWQPGTGPLPASAYII--- 235
Db 636 SVIE-----LPPPPSNVVERLPEPQQASINVSPTPGFDGNSPISKFIQRR 682
Qy 236 -----EAFSQSVSN---SWQT VANHVKT--TLYTVRGLRPNTIYLFMVRAIN----- 277
Db 683 EVSELEKFGVGPVDPDLLNWITELSNVSADQRWILLENKAAATVYQFRVSAVN RVGEGSPS 742
Qy 278 PKVSVTQXKPKNNGSTWAVNPLPP-----PPVQPLPGTELEHYAVE 319
Db 743 EPSNVVLPQEAHSG-----PPVGFVGSARSMSSEIITQWQPLEEHRNGQILGYILR 794
Qy 320 QQENGYSDSWCPPLPVQTYLHQGLEDELEED 351
Db 795 YRLFYNNVWPWS-----YQNITNEAQRN 817

RESULT 15
Q9IAJ0
ID Q9IAJ0 PRELIMINARY; PRT; 1788 AA.
AC Q9IAJ0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE RECEPTOR PROTEIN TYROSINE PHOSPHATASE LAR.
GN XTPP-LAR.
OS *Xenopus laevis* (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;

OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193505; PubMed=10727868;
RA Johnson K.G., Holt C.E.;
RT "Expression of CRYP-alpha, LAR, PTP-delta, and PTP-rho in the
RT developing xenopus visual system.";
RL Mech. Dev. 92:291-294(2000).
DR EMBL; AF197945; AAF43606.1; -.
SQ SEQUENCE 1788 AA; 200270 MW; AB192549866D9067 CRC64;

Query Match 14.5%; Score 330; DB 13; Length 1788;
Best Local Similarity 29.8%; Pred. No. 2.2e-16;
Matches 108; Conservative 47; Mismatches 154; Indels 54; Gaps 15;

Qy 1 QIVAQGRVTFPCETKGNPQPAVFWQKEGSQNLFPNQPPQNSRCSVPTGDLTITNIQ 60
: : | | | | | | | | | | : : | | : : | | : :
Db 141 KVVEKTRTATMLCAASGNPDPEITWFKD-----FLPVDTASSNGRIKQLRSGALQIENSE 195

Qy 61 RSDAGYYICQALTVAGS-ILAKAQLVTDVLTDRPPPIILQGPANQTLAVDGTALLKCKA 119
| | | | | | | | : | | | | | | | | : | : | : | :
Db 196 ESDQGGYECVATNSAGTRYSAANLYVR---VRRVAPRFSIPPSNHEVMPGGSVNLTCVA 252

Qy 120 TGDPLPVISWLKEGFTFPGDRPRATIQEQTLQIKNLRISDTGTYTCVATSSSGEASWSA 179
| : | : | : | : : : | : | : | : | | | | | |
Db 253 VGARMPYKWM-AGLEELTKEDEMPVGRNG-LELTN--IKDSANYTCVAISLGM--EA 306

Qy 180 VLDVTESGATISKNYDLSDLPGPPSPQVTDVTKNSVTLSWQPGTPTLPASAYIIIEAFS 239
| : | : : | | | | | : | | | | | | | | : :
Db 307 VAQIT-----VKALPKPPLDAMVTETTATSVTLTWDSGNPD--PVSYVVIQYKP 353

Qy 240 QSVNSWQTVANHVKTTLTYVRGLRPNTIYLFMVRAIN-----PKVSVTQXKPKXNGST 294
: : : | : | | | : | | : | : | : | : | : : : :
Db 354 KASESSFQEV-DGVATTRYSGGLSPFSEYEFRIIAVNNIGRGPPEVIEAQTGEQAPSS 412

Qy 295 WANVPLPPPPVQP---LPGTELEHYAVEQQENG-----YDSDSWCPPLPVQTYLHQG 343
| | | | | : : : | | : | | | :
Db 413 -----PPLKVQARMLASTMLVQWDLPEEPNGQIRGFRVYTTD---PHLPFSMWQKH 463

Qy 344 LED 346
: |
Db 464 VVD 466

Search completed: January 22, 2001, 12:54:12
Job time: 2053 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:19:47 ; Search time 233.01 Seconds
(without alignments)
21.719 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAVAAAAEYAGLKVARRQM.....REALDGRQVTLRTNPSPDR 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	761	100.0	148	20	Y13568	Mouse Robo 1 polyp
2	761	100.0	148	20	Y08406	Mouse partial ROBO
3	691	90.8	1649	20	Y08404	Human ROBO1 protei
4	691	90.8	1651	20	Y13566	Human Robo 1 polyp
5	108	14.2	434	20	Y13567	Human Robo 2 polyp
6	108	14.2	434	20	Y08405	Human partial ROBO
7	101	13.3	515	19	W72076	HSV-2 strain SB5 C
8	99.5	13.1	308	13	R22248	Sequence of rye-gr
9	97.5	12.8	308	20	Y25601	Lolium sp. allerge
10	97.5	12.8	414	19	W72159	HSV-2 strain SB5 C
11	97.5	12.8	414	19	W72139	HSV-2 strain SB5 C
12	88.5	11.6	821	14	R35451	Mouse eps8. Mus m

13	88	11.6	980	19	W69741	SAPAP2 protein. H
14	88	11.6	980	19	W69743	SAPAP1 protein. H
15	85	11.2	2424	21	Y78901	Calcium channel al
16	84	11.0	541	19	W37148	Mammalian Ena (Men
17	83	10.9	1319	14	R47043	Mammalian son of s
18	83	10.9	1336	16	R84638	mSOS1 protein. Mu
19	82	10.8	903	19	W77289	Zebrafish differen
20	81.5	10.7	449	19	W59866	Amino acid sequenc
21	81.5	10.7	456	19	W59865	Amino acid sequenc
22	81	10.6	1105	19	W44864	Human TPC2 telomer
23	81	10.6	1105	20	W73958	Human TPC2 protein
24	80.5	10.6	197	21	Y75526	Neisseria meningit
25	80.5	10.6	224	20	Y35778	Chlamydia pneumoni
26	80.5	10.6	548	18	W07700	Human ETS2 repress
27	80	10.5	127	18	W31265	Neospora caninum a
28	80	10.5	314	21	Y32321	Soybean beta-carot
29	80	10.5	386	20	Y59881	Human normal uteru
30	79.5	10.4	384	18	W23278	Bordetella pertuss
31	79.5	10.4	493	20	Y05396	Human TIE ligand N
32	79.5	10.4	493	21	Y70745	PSEQ-3 protein enc
33	79.5	10.4	493	21	Y54999	Human scarface 1 p
34	79	10.4	118	18	W19831	Plasmid pSP2alpha
35	79	10.4	256	20	Y00923	Alternatively spli
36	79	10.4	276	20	Y00922	Human CLAR1 protei
37	79	10.4	609	14	R32131	CMV Colburn region
38	78.5	10.3	350	20	Y35922	Extended human sec
39	78.5	10.3	543	18	W07702	Mouse ETS2 repress
40	78.5	10.3	783	19	W37151	Mouse neural Mena+
41	78.5	10.3	787	19	W37152	Mouse neural Mena+
42	78.5	10.3	802	19	W37153	Mouse neural Mena+
43	78.5	10.3	1784	16	R77223	Tuberous sclerosis
44	78	10.2	1333	21	Y68820	Amino acid sequenc
45	77.5	10.2	390	20	Y35923	Extended human sec

ALIGNMENTS

RESULT 1
Y13568
ID Y13568 standard; Protein; 148 AA.
XX
AC Y13568;
XX
DT 30-JUL-1999 (first entry)
XX
DE Mouse Robo 1 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Mus sp.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338003/28.
DR N-PSDB; X55772.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 50; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm

CC (commisureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

XX
SQ Sequence 148 AA;

Query Match 100.0%; Score 761; DB 20; Length 148;
Best Local Similarity 100.0%; Pred. No. 4.8e-71;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAVAAAAYAGLKVARRQMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAK 60
|
Db 1 aqavaaaayaglkvarrqmqdaagrrhfhasqcprptspvstdsnmsavviqkarpak 60

61 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVFKLASIEARTDRSSD 120
|
Db 61 qkhqpghlrreayaddlppppvpppaiksptvqskaqlevrpmvfklasieartdrssd 120

Qy 121 RKGGSYKGREALDGRQVTDLRNPSDPR 148
|
Db 121 rkgsykgrealdgrqvtdlrtnpsdpr 148

RESULT 2
Y08406
ID Y08406 standard; Protein; 148 AA.
XX
AC Y08406;
XX
DT 24-JUL-1999 (first entry)
XX
DE Mouse partial ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Mus sp.
XX
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
20-OCT-1998; 98WO-US22164.
XX
14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X57255.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 74-75; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX

SQ Sequence 148 AA;

Query Match 100.0%; Score 761; DB 20; Length 148;
Best Local Similarity 100.0%; Pred. No. 4.8e-71;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAVAAAAYAGLKVARRQMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAK 60
|
Db 1 aqavaaaayaglkvarrqmqdaagrrhfhasqcprptspvstdsnmsavviqkarpak 60

Qy 61 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVFKLASIEARTDRSSD 120
|
Db 61 qkhqpghlrreayaddlppppvpppaiksptvqskaqlevrpmvfklasieartdrssd 120

Qy 121 RKGGSYKGREALDGRQVTDLRNPSDPR 148
|
Db 121 rkgsykgrealdgrqvtdlrtnpsdpr 148

RESULT 3
Y08404
ID Y08404 standard; Protein; 1649 AA.
XX
AC Y08404;
XX
DT 24-JUL-1999 (first entry)
XX
DE Human ROBO1 protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
KW cell morphology; screening assay.
XX
OS Homo sapiens;
XX
PN WO9920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
PR 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI; 1999-312615/26.
DR N-PSDB; X08404.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 65-71; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
CC in nerve guidance which have been isolated from Drosophila sp.,
CC C. elegans, human and murine samples. The products of the invention can
CC be used to raise anti-Robo antibodies, which can be used to modulate cell
CC function or morphology. The Robo polynucleotides and fragments are useful
CC as probes and primers and for production of the Robo polypeptides. The
CC probes and primers are also useful in screening assays.
XX
SQ Sequence 1649 AA;

Query Match 90.8%; Score 691; DB 20; Length 1649;
Best Local Similarity 88.5%; Pred. No. 1.2e-62;
Matches 131; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 AQAVAAAAYAGLKVARRQMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAK 60
|
Db 1402 aqavaaaayaglkvarrqmqdaagrrhfhasqcprptspvstdsnmsaavmqktrpakk 1461

QY 61 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVPKLASIEARTDRSSD 120
||||||| | |||||:||||| :|||
Db 1462 lkhqpghlrretytdlppppvpppaiksptagsktqlevrpvvvpklpsmdartdrssd 1521

QY 121 RKGGSYKGREALDGRQVTDLTNPSPDR 148
||| ||||| ||||| :||| |||
Db 1522 rkgssykgrevldgrqvvdmrtnpgdpr 1549

RESULT 4
Y13566
ID Y13566 standard; Protein; 1651 AA.
XX
AC Y13566;
XX
DT 30-JUL-1999 (first entry)
Human Robo 1 polypeptide.
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
PN WO9925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55770.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 44-48; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 1651 AA;

Query Match 90.8%; Score 691; DB 20; Length 1651;
Best Local Similarity 88.5%; Pred. No. 1.2e-62;
Matches 131; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 AQAVAAAEYAGLKVARRQMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAKK 60
||||||| | |||||:||||| :|||
Db 1404 aqavaaaeyaglkvarrqmqdaagrhrhfhasqcprptspvstdsnmsaavmqktrpakk 1463

QY 61 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVPKLASIEARTDRSSD 120
||||||| | |||||:||||| :|||
Db 1464 lkhqpghlrretytdlppppvpppaiksptagsktqlevrpvvvpklpsmdartdrssd 1523

QY 121 RKGGSYKGREALDGRQVTDLTNPSPDR 148
||| ||||| ||||| :||| |||
Db 1524 rkgssykgrevldgrqvvdmrtnpgdpr 1551

RESULT 5
Y13567
ID Y13567 standard; Protein; 434 AA.
XX
AC Y13567;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
KW modulation; nerve cell function.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 285
FT /label= unknown
FT /note= "encoded by GTN"
FT Misc-difference 396
FT /label= unknown
FT /note= "encoded by NTT"
XX
PN WO9925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI; 1999-338008/28.
DR N-PSDB; X55771.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure; Page 49-50; 56pp; English.
XX
CC The invention relates to a method for modulating the amount of Comm
CC (commissureless) polypeptide in contact with a cell expressing active
CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.
XX
SQ Sequence 434 AA;

Query Match 14.2%; Score 108; DB 20; Length 434;
Best Local Similarity 57.9%; Pred. No. 0.0023;
Matches 22; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 29 PHASQCPRPTSPVSTDSNMSAVVIQKARPAKKQKQPG 66
|:| | ||||| ||||| | :| | | :||:|
Db 397 ftssqprptspfstdsntsaalsqsrprptkhhkgg 434

RESULT 6
Y08405
ID Y08405 standard; Protein; 434 AA.
XX
AC Y08405;
XX

```

RESULT      8
R22248
ID   R22248 standard; Protein; 308 AA.
XX
AC   R22248;
XX
DT   22-JUL-1992 (first entry)
XX
DE   Sequence of rye-grass pollen allergens en
DE   lambda-12R.
XX
KW   Rye grass pollinosis; diagnosis; therapy.
XX
OS   Lolium perenne.
XX
FH   Key                      Location/Qualifiers
FT   Peptide                  1..25
XX                               /label= signal
XX
PN   W09203550-A.
XX
PD   05-MAR-1992.
XX
PF   16-AUG-1991;           91WO-AU00369.
XX
PR   17-AUG-1990;           90AU-0001823.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Singh MB, Hough T, Knox RB, Avjioglu A
XX
DR   WPI; 1992-096894/12.
DR   N-PSDB; 023000.

```

XX
PT New nucleic acid sequences coding rye-grass pollen allergens -
PT esp. Lol pIa and Lol pIb and their fragments, for diagnosing and
PT detecting rye-grass pollinosis
XX
PS Disclosure; ; 81pp; English.
XX
CC The inventors claim a sequence encoding the rye grass pollen
CC allergen Lol pIa, or an antigenic fragment. The allergen can
CC alternatively be Lol pIb. The antigenic fragment has T-cell
CC stimulating activity and IgE stimulating activity. It does not bind
CC IgE specific for rye grass pollen however. It may be encoded by
CC clone 12R (Q23000) or 26.j (Q22246). R22248 is as printed in the
CC specification; several codons in Q23000 do not translate into the
CC amino acids written in the specification.
XX
SQ Sequence 308 AA;
Query Match 13.1%; Score 99.5; DB 13; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.011;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;

Qy 3 AVAAAAEYAGLKVARRQMQDAAGRHHFASQCPRPTSPVSTDSNMSAVVIQKARPAKKQK 62
| | | : : : || | || : | |
Db 39 atpaatpaggwregddrraeaggrgrlasrqpwplptp----- 78
Qy 63 HQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVVKLASIEARTDRSSDRK 122
||| : || | || | || : || | : || | : | |
Db 79 -----lrrtssststpspsppra-spsaaka-----pglipkl-----dtayd-- 118
Qy 123 GGSYKGREALDGRQVTLRTNP 144
: || | | | | | |
Db 119 -vaykaeahprgqvrrlrhpc 139

RESULT 9
Y25601
ID Y25601 standard; protein; 308 AA.
XX
AC Y25601;
XX
DT 30-SEP-1999 (first entry)
XX
DE Lolium sp. allergen 2498581 Lol p 5a protein fragment.
XX
KW Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
KW chiromidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.
XX
OS Lolium sp.
XX
PN WO9934826-A1.
XX
PD 15-JUL-1999.
XX
PF 11-JAN-1999; 99WO-GB00080.
XX
PR 21-SEP-1998; 98GB-0020474.
PR 09-JAN-1998; 98GB-0000445.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Kay AB, Larche M;
XX
DR WPI; 1999-458255/38.
XX
PT Desensitizing patients to polypeptide allergens
XX
PS Example 6; Page 55; 117pp; English.

XX
CC This invention describes a novel method of desensitizing a patient to a
CC polypeptide allergen and comprises administering to the patient a peptide
CC derived from the allergen where restriction to a MHC Class II molecule
CC possessed by the patient can be demonstrated for the peptide and the
CC peptide is able to induce a late phase response in an individual who
CC possesses the MHC Class II molecule. The methods can be used for
CC desensitising patients to allergens present in e.g. grass, tree and weed
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,
CC the chiromidae (non-biting midges), spiders and mites, housefly, fruit
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of
CC Tenbrio molitor beetle, mammals such as cat, dog, horse, cow, pig,
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to
CC produce immunological vaccines which may be used to prevent and/or treat
CC conditions involving hypersensitivity to allergens. This sequence
CC represents the Lolium sp. allergen 2498581 Lol p 5a.
XX
SQ Sequence 308 AA;

Query Match 12.8%; Score 97.5; DB 20; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.018;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;

Qy 3 AVAAAAEYAGLKVARRQMQDAAGRHHFASQCPRPTSPVSTDSNMSAVVIQKARPAKKQK 62
| | | : : : || | || : | |
Db 39 atpaatpaggwregddrraeaggrgrlasrqpwplptp----- 78
Qy 63 HQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPVMVVKLASIEARTDRSSDRK 122
||| : || | || | || : || | : || | : | |
Db 79 -----lrrtsssrpspsppra-spsaaka-----pglipkl-----dtayd-- 118
Qy 123 GGSYKGREALDGRQVTLRTNP 144
: || | | | | | |
Db 119 -vaykaeahprgqvrrlrhpc 139

RESULT 10
W72159
ID W72159 standard; Protein; 414 AA.
XX
AC W72159;
XX
DT 08-JAN-1999 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 12 ORF#1 protein.
XX
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
XX
OS Herpes simplex virus type 2.
XX
PN WO9820016-A1.
XX
PD 14-MAY-1998.
XX
PF 31-OCT-1997; 97WO-US20016.
XX
PR 09-JUN-1997; 97US-0049018.
PR 04-NOV-1996; 96US-0030279.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI Esser KM, Leary JJ;
XX
DR WPI; 1998-286847/25.
DR N-PSDB; V62175.
XX
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal

PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
PS Claim 10; Page 97-98; 748pp; English.
XX
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 3.
CC Based on homology, this sequence is a immediate-early protein IE68.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,

Qy 120 DRKGGSYKGREALDGRQVTDLR-----TNPSD 146
| :|| : || || | ||
Db 63 dlraqrrrrsprelaaryfldlsaettqtateseqatpsd 100

Qy 2 QAVAAAAEAYAGLKVARRQQMDAAGRHHFHASQCPRPTSPVSTDSNMSAVVIQKARPAKKQ 61
||| | : : | : : | || : : : ||| | ||
Db 137 qavvhacsydsi-lalvckeptgskpdlhlfqcdvkanlisediesaisdsk---ggyk 192

Qy 62 KHOPGHLRRELYAD-DLPPP-----PVPPPAIKSPTVQSKA----- 96

Db 193 krrpealmlakadpgippprapapvppgtvtqvdvrsrvaawsawaadqgdfekprqy 252
Qy 97 -QLEVRPM-----VPKL-ASIEARTDRSSDRKGSYKREAL 132
Db 253 hegeetpemaaridrdvqilnhildieffitklgkaaeafelskrkk--skkskrkg 310
Qy 133 DGRQVTDLRNPSDP 147
Db 311 pgegvitlrakpppp 325

RESULT 13
W69741
ID W69741 standard; protein; 980 AA.
XX
AC W69741;
26-OCT-1998 (first entry)
DE SAPAP2 protein.
XX
KW Human; SAPAP1; SAPAP2; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
KW membrane associated guanylate kinase; neuronal disease.
XX
OS Homo sapiens.
XX
PN JP10201477-A.
XX
PD 04-AUG-1998.
XX
PF 24-JAN-1997; 97JP-0011714.
XX
PR 24-JAN-1997; 97JP-0011714.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKE/) TAKEUCHI M.
XX
DR WPI; 1998-474491/41.
XX
PT New protein SAPAP1 - used for, e.g. diagnosis and prevention of
PT various neuronal diseases
XX
PS Disclosure; Page 7-9; 12pp; Japanese.
XX
The present sequence represents the SAPAP2 protein which is mentioned in
the present invention. The present invention specifically claims the
SAPAP1 protein which has a 992 amino acid (aa) sequence. Also described
in the present invention are: (1) an animal protein having an aa sequence
substantially homologous to SAPAP1; (2) cDNA sequence encoding SAPAP1,
or an aa sequence substantially homologous to SAPAP1, and (3) a genomic
DNA sequence hybridised to the cDNA or its partial sequence. SAPAP1 is a
novel animal protein specific for PSD-95/SAP90 and its related protein,
and may be useful for the diagnosis, prevention and treatment of various
neuronal diseases caused by functional or structural interference of
nervous system.
XX
SQ Sequence 980 AA;

Query Match 11.6%; Score 88; DB 19; Length 980;
Best Local Similarity 26.6%; Pred. No. 0.66;
Matches 29; Conservative 13; Mismatches 39; Indels 28; Gaps 4;

Qy 40 PVSTDSNMSAVVIQKARPA---KKQKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSK 95
Db 512 pvmtpsnmtstirstaavsytnykk-----tppvpvprttskplisvt 554
Qy 96 AQLEVRPMVMPKASIEARTDRSS---DRKGSYKREALDGRQVTDL 140
Db 555 aqsstest---qdayqdsraqrmspwpqdsrgglynsmdsldsnkamnl 600

RESULT 14
W69743
ID W69743 standard; protein; 980 AA.
XX
AC W69743;
XX
DT 26-OCT-1998 (first entry)
XX
DE SAPAP1 protein.
XX
KW Human; SAPAP2; SAPAP1; animal protein; PSD-95/SAP90; diagnosis;
KW nervous disease; functional interference; structural interference;
KW membrane associated guanylate kinase; neuronal disease.
XX
OS Homo sapiens.
XX
PN JP10201478-A.
XX
PD 04-AUG-1998.
XX
PF 24-JAN-1997; 97JP-0011715.
XX
PR 24-JAN-1997; 97JP-0011715.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (TAKE/) TAKEUCHI M.
XX
DR WPI; 1998-474492/41.
XX
PT DNA encoding new animal protein SAPAP 2 - useful for diagnosis and
PT treatment of nervous system diseases
XX
PS Disclosure; Page 4-6; 11pp; Japanese.
XX
CC The present sequence represents the SAPAP1 protein, which is mentioned
CC in the present invention which specifically claims SAPAP2. Also described
CC in the present invention is: (A) an animal protein having an amino acid
CC sequence substantially the same as SAPAP2; (B) a cDNA sequence encoding
CC the amino acid sequence of SAPAP2 or (C) an amino acid sequence
CC substantially same as SAPAP2; and (D) a genomic DNA sequence hybridised
CC by the above cDNA or its partial sequence. SAPAP2 is a new animal
CC protein which combines specifically with PSD-95/SAP90 and its related
CC protein and is useful for the diagnosis, prevention and treatment of
CC various nervous diseases caused by functional or structural interference
CC of nervous system.
XX
SQ Sequence 980 AA;

Query Match 11.6%; Score 88; DB 19; Length 980;
Best Local Similarity 26.6%; Pred. No. 0.66;
Matches 29; Conservative 13; Mismatches 39; Indels 28; Gaps 4;

Qy 40 PVSTDSNMSAVVIQKARPA---KKQKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSK 95
Db 512 pvmtpsnmtstirstaavsytnykk-----tppvpvprttskplisvt 554
Qy 96 AQLEVRPMVMPKASIEARTDRSS---DRKGSYKREALDGRQVTDL 140
Db 555 aqsstest---qdayqdsraqrmspwpqdsrgglynsmdsldsnkamnl 600

RESULT 15
Y78901
ID Y78901 standard; protein; 2424 AA.
XX
AC Y78901;
XX
DT 19-MAY-2000 (first entry)
XX
DE Calcium channel-alpha 1A amino acid sequence.
XX

KW Alpha 1A subunit; central nervous system calcium channel; dementia;
KW Alzheimer's disease; calcium channel function evaluation;
KW cerebral function.
XX
OS Xenopus sp.
XX
PN JP2000026315-A.
XX
PD 25-JAN-2000.
XX
PF 09-JUL-1998; 98JP-0194236.
XX
PR 09-JUL-1998; 98JP-0194236.
XX
PA (DAUC) DAIICHI PHARM CO LTD.
XX
DR WPI; 2000-176900/16.
XX
PT Evaluating calcium channel activator - involves measuring inhibition or
PT change in functions of coupling between alpha subunit of G-protein and
PT alpha-1 subunit of C-terminal area, used for treating dementia and
T improving cerebral functions
X
PS Example 1; Page 9-10; 11pp; Japanese.
XX
CC This sequence represents the alpha 1A subunit of the Xenopus calcium
CC channel of the central nervous system. The invention relates to the
CC evaluation of a calcium channel activator. The evaluation process
CC consists of measuring the inhibition or change in function of coupling
CC between a G-protein alpha subunit and the alpha 1 subunit of the calcium
CC channel of the central nervous system. Test compounds which are found to
CC cause the inhibition of coupling or a change in function can be
CC identified from tests carried out on genetically engineered cells. The
CC evaluation method may be used to discover compounds for treating dementia
CC and improving cerebral function in diseases such as Alzheimer's disease.
XX
SQ Sequence 2424 AA;

Query Match 11.2%; Score 85; DB 21; Length 2424;
Best Local Similarity 25.6%; Pred. No. 3.9;
Matches 34; Conservative 15; Mismatches 46; Indels 38; Gaps 6;

Qy 18 RQMDAAGRRHFHASQCPRPTSPVSTDS---NMSAVVIQKARPAKQKHQPG----- 66
| : : | | : | | | : | | | : : | | | | : | |
Db 2294 rrrrgggggra---lrrapggprelaqdsppgrgsvclaraarpagpqrllpgprtggqapr 2350

Qy 67 -----HLRREAYADDLPPPPVPPPAIKSPTVQSKAQLVPRVPMVKLASIEARTDR 117
: | | | | | | | : | | : | : | : |
b 2351 arlpqkparsvqrrrrglvlsppp-pppgelap-----rahpartprpgpgdsrsrr 2401

y 118 SSDR-----KGG 124
| | |
Db 2402 ggrrwtasagkg 2414

Search completed: January 22, 2001, 12:19:50
Job time: 1747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:27:28 ; Search time 325.28 Seconds
(without alignments)
30.894 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAVAAAAEYAGLKVARRQM.....REALDGRQVTDLRTPNSDPR 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	761	100.0	1612	2	T30805	duttl protein - mo
2	719	94.5	1651	2	T14160	transmembrane rece
3	101	13.3	1344	2	T14316	rig-1 protein - mo
4	99	13.0	739	2	T21431	hypothetical prote
5	97.5	12.8	308	2	A38582	pollen allergen pI
6	91	12.0	825	2	T27852	hypothetical prote
7	89	11.7	1882	1	GNVVTR	genome polypeptide
8	88.5	11.6	821	2	S39983	eps8 protein - mou
9	87	11.4	826	2	A60385	monocyte surface a
10	86.5	11.4	891	2	T22560	hypothetical prote
11	85.5	11.2	662	2	T23757	hypothetical prote
12	85.5	11.2	1870	2	S37671	MHC class III hist
13	85.5	11.2	1872	2	S36152	MHC class III hist
14	85	11.2	300	2	T49225	hypothetical prote
15	85	11.2	698	2	T32594	hypothetical prote
16	85	11.2	716	2	T26998	hypothetical prote
17	85	11.2	1541	2	T02831	AAA protein L4171.
18	85	11.2	2424	2	I46480	calcium channel BI
19	85	11.2	2424	2	I46479	calcium channel BI
20	83.5	11.0	282	2	S53502	histone H1 - commo
21	83.5	11.0	365	2	T24955	hypothetical prote
22	83	10.9	1336	2	S25716	Ras guanine nucleo
23	83	10.9	1560	2	T42727	proliferation pote
24	82.5	10.8	1013	2	T46422	hypothetical prote
25	82.5	10.8	1102	2	JC6316	probable protein k
26	82.5	10.8	1547	2	T28657	blackjack protein,
27	82	10.8	363	2	T16755	hypothetical prote
28	82	10.8	601	2	S56144	SH3 domain binding
29	82	10.8	797	1	I46044	furin (EC 3.4.21.7

30	82	10.8	867	2	JH0225	L96 protein - Tipu
31	81.5	10.7	537	2	A46611	myosin-binding pro
32	81.5	10.7	555	2	I53869	zinc finger protei
33	81.5	10.7	555	2	T30349	structural protein
34	81.5	10.7	806	2	T35640	probable sensor ki
35	81.5	10.7	822	2	I38728	epidermal growth f
36	81.5	10.7	963	2	T04002	hypothetical prote
37	81.5	10.7	975	2	T08606	protein phosphatas
38	81.5	10.7	1603	2	S17983	gene posterior sex
39	80.5	10.6	197	2	F82029	probable periplasm
40	80.5	10.6	417	2	A72078	ct005 hypothetical
41	80.5	10.6	417	2	B81590	conserved hypothet
42	80.5	10.6	548	2	S59133	ETS2 repressor fac
43	80.5	10.6	2142	2	B35098	MHC class III hist
44	80	10.5	202	2	T11744	dehydrin - kidney
45	80	10.5	383	1	A48222	dematin 48K chain

ALIGNMENTS

RESULT 1
T30805
duttl protein - mouse
N;Alternate names: transmembrane receptor protein Robol homolog
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30805
R;Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.
submitted to the EMBL Data Library, July 1998
A;Description: The mouse homologue of human DUTTL/H-robol gene: protein sequence and
A;Reference number: Z20879
A;Accession: T30805
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-1612 <WUM>
A;Cross-references: EMBL:Y17793; NID:e1329712; PID:e1329713; PIDN:CAA76850.1
A;Experimental source: brain
C;Genetics:
A;Gene: duttl
A;Map position: 16

Query Match 100.0%; Score 761; DB 2; Length 1612;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAVAAAAEYAGLKVARRQMADAAGRRHFASQCPRPTSPVSTDSNMSAVVIQKRAPAKK 60
|||||
Db 1365 AQAVAAAAEYAGLKVARRQMADAAGRRHFASQCPRPTSPVSTDSNMSAVVIQKRAPAKK 1424
Qy 61 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPMVVKLASIEARTDRSSD 120
|||||
Db 1425 QKHQPGHLRREAYADDLPPPPVPPPAIKSPTVQSKAQLEVRPMVVKLASIEARTDRSSD 1484
Qy 121 RKGGSYKGREALDGRQVTDLRTPNSDPR 148
|||||
Db 1485 RKGGSYKGREALDGRQVTDLRTPNSDPR 1512

RESULT 2
T14160
transmembrane receptor protein Robol - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14160
R;Kidd, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodman
Cell 92, 205-215, 1998
A;Title: Roundabout controls axon crossing of the CNS midline and defines a novel class
A;Reference number: Z17897; MUID:98117249
A;Accession: T14160
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA

RESULT 6
T27852
hypothetical protein ZK418.6 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 15-Oct-1999

```
Qy      2 QAVAAAAEYAGLKVARROMQDAAGR---RHFASQCPRPTSPVSTDSNMSAVVIQKARPA 58
       :| || | | : | | | | : | | | : | | |
Db    193 KAAALAAVKAAQEAAPRLAAQKAAISKILRDROVAALPPPPPSAARLAAEAEASKAESL 252
       :| || | | : | | | | | | | | | | | | | | | | | | | | | |
Qy     59 KKQKHQPCHLR-REAYADDLP PPPVP PPAIKSPTVQSKAQLEVRPVMVPKLASIEARTDR 117
       :: | | | | | | | | | | | | | | | | | | | | | |
Db    253 RRLKAFTFSFVRPALNTSFPPPPPPPPARSSEL-----LAAFEAMNR 296
       :| | | | | | | | | | | | | | | | | | | | | |
Qy    118 SSDRKGG 124
       | :||
Db    297 SOPVOGG 303
```

Qy 4 VAAAEYAGLVARRQQMDAAGRHHFASQCPRPTSPVST-----DSNMSAVVIQKARP 57
 ||| |::: |: | | :|: | |::: | |
Db 671 VAAMVIVAGIYLR-----APRQIQRRSVAPRKISGLSNPLFTTRDSSL-----P 716

Qy 58 AKKQKHQPCHLREAYADDLPVPVP-----PPPAIKSPTVQSQAQLEVRPMVMVKLASIE 112
 ||| |::: | | | | | |

Db 717 AKNRPPDPS----ETVSTNQPPRIAPKPRPPA--PPGAVSSSLPV-PVYAPKIPN-Q 768

Qy 113 ARTDRSSDRKGSYKRGREALDGRQVTDLRINPSDP 147

Db 769 FRPDPT-----KPLPELKPKQVKPTFAPPTPP 796

RESULT 10

T22560

hypothetical protein F53C11.5 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T22560

R;Baynes, C.

submitted to the EMBL Data Library, September 1996

A;Reference number: Z19581

A;Accession: T22560

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-891 <WIL>

A;Cross-references: EMBL:Z79756; PIDN:CAB02122.1; GSPDB:GN00023; CESP:F53C11.5

A;Experimental source: clone F53C11

A;Genetics:

A;Gene: CESP:F53C11.5

A;Map position: 5

A;Introns: 25/3; 59/1; 137/1; 287/1; 313/1; 343/3; 420/1; 455/2; 637/2; 708/2; 831/2; 86

Query Match 11.4%; Score 86.5; DB 2; Length 891;
Best Local Similarity 29.0%; Pred. No. 7.1;
Matches 29; Conservative 8; Mismatches 40; Indels 23; Gaps 3;

Qy 32 SQCPRTSPVST-----DSNMSAVVIQKARPAKKQKHQPGHLRREAYADLPPPPVPP 84

Db 463 SQSPRPSQPLQTIAPPPSSQSSNIPPSLPSPRPSALP--APGFSRQISSATTLTAQAPP 520

Qy 85 PAIKSPTVQSKAQLEVRPVMVPKLASIEARTDRSSDRKGG 124

Db 521 QALTSPRTPS-----TSASPTSSRFRDSGG 546

RESULT 11

T23757

hypothetical protein M117.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T23757

R;Kershaw, J.

submitted to the EMBL Data Library, June 1996

A;Reference number: Z19794

A;Accession: T23757

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-662 <WIL>

A;Cross-references: EMBL:Z73910; PIDN:CAA98136.1; GSPDB:GN00022; CESP:M117.4

A;Experimental source: clone M117

A;Genetics:

A;Gene: CESP:M117.4

A;Map position: 4

A;Introns: 19/2; 136/3; 281/3; 399/2; 556/3; 602/2

Query Match 11.2%; Score 85.5; DB 2; Length 662;
Best Local Similarity 25.4%; Pred. No. 6.3;
Matches 32; Conservative 17; Mismatches 58; Indels 19; Gaps 4;

Qy 30 HASQCPRTSPVSTDSNMSAVVIQKARPAKKQKHQPGHLRREAYADLPPPPVPPPAIKS 89

Db 126 HAHVSPDPTSLNDKSKFEKPNLEPGKPEKPEKLD-----KPLAKPLPPSAAPPV-A 178

Qy 90 PTVQSKAQLEVRPVMVP-----KLASIEARTDRSSDRKGSYKRGREALDGRQVTDLR 142

Db 179 PLSAPVSTGAPPSAVPPADKKMSQSTRDKKSGRSTRDNIHKKDKVAALD-----ENKK 233

Qy 143 NPSPDR 148

Db 234 SPEDKK 239

RESULT 12

S37671

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu

C;Species: Homo sapiens (man)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000

C;Accession: S37671

R;Bougueleret, L.

submitted to the EMBL Data Library, August 1992

A;Reference number: S37671

A;Accession: S37671

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1870 <BOU>

A;Cross-references: EMBL:Z15025; NID:g29374; PID:g29375

C;Genetics:

A;Map position: 6p21.3

A;Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;

C;Superfamily: collagen alpha 1(IV) chain

Query Match 11.2%; Score 85.5; DB 2; Length 1870;
Best Local Similarity 25.7%; Pred. No. 19;
Matches 39; Conservative 14; Mismatches 58; Indels 41; Gaps 7;

Qy 18 RQMQDAAGRRHFHASQCPRTSPVSTDS-NMSAVVIQKARPAKKQKHQPGHLRREAYADD 76

Db 656 KQQQHQHQHQHQHQGSAPPTVPVPPSPQPVTLGAVPAPKAPPPPPKALYPGALGR----- 709

Qy 77 LPPPPVP-----PPAIKSPTVQSKAQLEVRPV-----MVPKLASIEARTDRSS 119

Db 710 --PPPPMPMNFDPWMMIPVYDPRLLQGRPLDFYPPGVHPSGLVPR-----ERS 758

Qy 120 DRKGSYKRGREALDGRQVTDLR---TNPSDP 148

Db 759 DSRGLS---SEPFDRHAPAMLRERGTTPVDPK 787

RESULT 13

S36152

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu

C;Species: Homo sapiens (man)

C;Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000

C;Accession: S36152

R;Iris, F.J.M.; Bougueleret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Ju

Nature Genet. 3, 137-145, 1993

A;Title: Dense Alu clustering and a potential new member of the NFkappaB family withi

A;Reference number: S36152; MUID:93272029

A;Accession: S36152

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1872 <IRI>

A;Cross-references: EMBL:Z15025

A;Note: in the authors' translation residues 32-34 are shown after residue 4 and, con

A;Note: the authors translated the codon AAT for residue 1000 as His

C;Genetics:

A;Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;

C;Superfamily: collagen alpha 1(IV) chain

Query Match 11.2%; Score 85.5; DB 2; Length 1872;
Best Local Similarity 25.7%; Pred. No. 19;
Matches 39; Conservative 14; Mismatches 58; Indels 41; Gaps 7;

Qy 18 RQMQDAAGRRHFHASQCPRTSPVSTDS-NMSAVVIQKARPAKKQKHQPGHLRREAYADD 76

Db 657 KQQQHQHQHQHQHQGSAPPTVPVPPSPQPVTLGAVPAPKAPPPPPKALYPGALGR----- 710

Qy 120 DRKGGSYKGREALDGRQVTDLR---TNPSDPR 148
| : | | | | || | | :
Db 760 DSRGLS---SEPFDRHAPAMLRERGTTPVDPK 788

RESULT 14

T49225

hypothetical protein F27H5.90 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C;Accession: T49225

R; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, R. submitted to the Protein Sequence Database, April 2000

Reference number: 225018

Accession: T49225

4/ Status: preliminary

A; Molecule type: DNA

A;Residues: 1-300 <RIE>

A;Cross-references: EMBL:AL163852; GSPDB:GN00061; ATSP:F27H5.90

A:Experimental source: cultivar Columbia; BAC clone F27H5

C: Genetics:

A;Gene: ATSP:F27H5.90

A:Map position: 3

A; Introns: 46/3; 59/3; 105/3; 127/1; 159/3

Query Match 11.2%; Score 85; DB 2; Length 300;
Best Local Similarity 25.3%; Pred. No. 2.9;
Matches 40; Conservative 23; Mismatches 51; Indels 44; Gaps 8;

Qy 26 RRHFHASQCPRPTSPVSTDSNM--SAVVIQKARPAKQ----KHQPGHLRREAYADDL-- 77
|: ||:| | |: |: | | | :||: |
Db 137 RRIFHSSDIEHVLVLGGAOSSLOPSSLI-----PGKDDPLQSDSNIIRREFESILKT 191

Qy 78 -----PPPPVPPPAKSPT-----VQSKAQLEVRPMVMPKIASIEA 113
| :||| : : :| : ||| |
Db 192 OEEKGLVPQKNISVVPGLPPPAHASTSRNEEEEGAGSSEOGEEE----PKPAESET 247

Qy 114 RTDRSSDRKG-GSYKGREALDGR--QVTDLRTNPSDPR 148
: ||:| | :|| || | : : | ||
Db 248 NSSSSTNRRGRGRWRGRGRSRGRGPTVNERKPNSDPR 285

SULT 15

T32594

hypothetical protein C02B10.5 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence revision 29-Oct-1999 #text change 29-Oct-1999

C;Accession: T32594

R; Nelson, J.; White, S.; Hawkins, J.; Wohldmann, P.

submitted to the EMBL Data Library, December 1997

A;Description: The sequence of *C. elegans* cosmid C02B10.

A;Reference number: 221196

A; Accession: T32594

A:Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-698 <NEL>

A;Cross-references: EMBL:AF038605; PIDN:AAB92020.1; GSPDB:GN00022; CESP:C02B10.5

A; Experimental source: strain Bristol N2; clone C02B10

C; Genetics:

A;Gene: CESP:C02B10.5

A:Map position: 4

A: Introns: 61/3: 102/2: 188/3: 349/2: 641/1

Query Match 11.2%; Score 85; DB 2; Length 698;
Best Local Similarity 31.0%; Pred. No. 7.3;
Matches 35; Conservative 10; Mismatches 48; Indels 20; Gaps 5;

Qy 1 AQAVAAAAAYAGLKVARR--QMQDAAGRRHFASQCPRPTSPVSTDSNMSAVVIQKARPA 58
|| ||| ||: | | : | | | | | | : | | |
Db 514 AOKAAAAAAGMGSPARASPMGASPSPPHPSPSP-PNHPANP---MYHHHMMWRAM 569

Qy 59 KKQKHQPGLRREAYADLPPPPVPPPAIKSPTVQSKAQLEVRPMVPKLASI 111
| | | | : | | : | :
Db 570 HAQGGO PGH-----PGMMPGMMPPGMPPGM-----MPPGMHPGMAGM 608

Search completed: January 22, 2001, 12:27:33

Job time: 2130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:29:51 ; Search time 162.41 Seconds
(without alignments)
29.429 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAVAAAAEYAGLKVARROM.....REALDGRQVTLRLNPSPDR 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB ID	Description
1	97.5	12.8 308 1 MP5A_LOLPR	Q40240 lolium pere
2	89	11.7 1882 1 POL2_TRSVR	P25247 tomato ring
3	88.5	11.6 821 1 EPS8_MOUSE	Q08509 mus musculus
4	86	11.3 826 1 AD08_MOUSE	Q05910 mus musculus
5	85	11.2 364 1 F812_HUMAN	P23610 homo sapien
6	85	11.2 2424 1 CCAA_RABIT	P27884 oryctolagus
7	84.5	11.1 197 1 IE68_HSV2	P14379 herpes simp
8	83	10.9 1319 1 SOS1_MOUSE	Q62245 mus musculus
9	82	10.8 601 1 3BP1_MOUSE	P55194 mus musculus
10	82	10.8 797 1 FUR1_BOVIN	Q28193 bos taurus
11	82	10.8 867 1 VL96_IRV1	P22856 tipula irid
12	81.5	10.7 537 1 MYPH_CHICK	Q05623 gallus gall
13	81.5	10.7 822 1 EPS8_HUMAN	Q12929 homo sapien
14	81.5	10.7 1603 1 PSC_DROME	P35820 drosophila
15	80.5	10.6 548 1 ERF_HUMAN	P50548 homo sapien
16	80.5	10.6 2142 1 BAT2_HUMAN	P48634 homo sapien
17	80	10.5 127 1 GR14_NEOCA	Q25540 neospora ca
18	80	10.5 383 1 DEMA_HUMAN	Q08495 homo sapien
19	80	10.5 2004 1 MO2_HUMAN	Q92794 homo sapien
20	79.5	10.4 415 1 ACRO_PIG	P08001 sus scrofa
21	79	10.4 136 1 SR19_ORYSA	P49964 oryza sativ
22	79	10.4 299 1 RL22_DROME	P50887 drosophila
23	79	10.4 484 1 OAR2_LOCFI	Q25322 locusta mig
24	79	10.4 589 1 VP40_SCMVC	P16046 simian cyto
25	78.5	10.3 180 1 US10_VZVD	P09311 varicella-z
26	78.5	10.3 190 1 PP28_HCMVA	P13200 human cytom
27	78.5	10.3 304 1 INO2_YEAST	P26798 saccharomyc
28	78.5	10.3 481 1 CAP_CHLVR	P40122 chlorohydra
29	78.5	10.3 515 1 HSMH_DROME	Q03372 drosophila
30	78.5	10.3 551 1 ERF_MOUSE	P70459 mus musculus
31	78	10.2 1333 1 SOS1_HUMAN	Q07889 homo sapien
32	77.5	10.2 309 1 J1L_HCMVA	P17143 human cytom
33	77.5	10.2 1125 1 MAP4_MOUSE	P27546 mus musculus

34	77	10.1	314	1	FTSQ_MYCTU	O06226 mycobacteri
35	77	10.1	448	1	TAU1_BOVIN	P29172 bos taurus
36	77	10.1	559	1	3BP2_MOUSE	Q06649 mus musculus
37	77	10.1	935	1	RNE_HAEIN	P44443 haemophilus
38	77	10.1	1080	1	MI15_CAEEL	Q23356 caenorhabdi
39	77	10.1	1394	1	CNG4_BOVIN	Q28181 bos taurus
40	76.5	10.1	251	1	HXB4_FUGRU	O13074 fugu rubrip
41	76.5	10.1	578	1	PSP2_YEAST	P50109 saccharomyc
42	76.5	10.1	678	1	ABPP_RIPCL	Q27905 riptortus c
43	76.5	10.1	1132	1	BAT3_HUMAN	P46379 homo sapien
44	76	10.0	484	1	OAR1_LOCFI	Q25321 locusta mig
45	76	10.0	520	1	RXRB_MOUSE	P28704 mus musculus

ALIGNMENTS

RESULT 1
MP5A_LOLPR
ID MP5A_LOLPR STANDARD; PRT; 308 AA.
AC Q40240;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAJOR POLLEN ALLERGEN LOL P 5A PRECURSOR (LOL P VA) (LOL P IB).
GN LOL PIB.
OS Lolium perenne (Perennial ryegrass).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Lolium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91142177; PubMed-1671715;
RA Singh M.B., Hough T., Theerakulpisut P., Avjioglu A., Davies S.,
RA Smith P.M., Taylor P., Simpson R.J., Ward L.D., McCluskey J.,
RA Puy R., Knox R.B.;
RT "Isolation of cDNA encoding a newly identified major allergenic
RT protein of rye-grass pollen: intracellular targeting to the
RT amyloplast.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:1384-1388(1991).
CC -!- SUBCELLULAR LOCATION: STARCH GRANULE.
CC -!- TISSUE SPECIFICITY: POLLEN, STARCH GRANULES.
CC -!- DISEASE: CAUSES GRASS POLLEN ALLERGY.
CC -!- SIMILARITY: BELONGS TO THE POA P IX/PHL P VI FAMILY OF ALLERGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M59163; AAA33406.1; -.
DR INTERPRO; IPR002914; -.
DR PFAM; PF01620; Pollen_allerg_2; 1.
KW Signal; Allergen; Multigene family.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 308 MAJOR POLLEN ALLERGEN LOL P 5A.
FT DOMAIN 31 46 ALA/PRO/THR-RICH.
FT DOMAIN 270 289 ALA/THR-RICH.
FT DOMAIN 33 36 POLY-ALA.
FT DOMAIN 270 278 POLY-ALA.
SQ SEQUENCE 308 AA; 31881 MW; 7756025D09E12FFF CRC64;

Query Match 12.8%; Score 97.5; DB 1; Length 308;
Best Local Similarity 29.6%; Pred. No. 0.25;
Matches 42; Conservative 10; Mismatches 49; Indels 41; Gaps 5;

Qy 3 AVAAAAEYAGLKVARROMQDAARRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAKKQK 62
| | | : : | | | : | | |
Db 39 ATPAATPAGGWRGDDRRRAEAGGRQLASRQWPPLPTP----- 78

Qy 133 DGRQVTDLRTNPSDP 147
| | || . | |
Db 311 PEGEVLTLRAKPPPP 325

```

RESULT 5
F8I2_HUMAN
ID  F8I2_HUMAN      STANDARD;      PRT;   364 AA.
AC  P23610;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE  FACTOR VIII INTRON 22 PROTEIN (CPG ISLAND PROTEIN).
GN  F8A.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=LIVER;
RX  MEDLINE=90243242; PubMed=2110545;
RA  Levinson B., Kenwright S., Lakich D., Hammonds G. Jr., Gitschier J.;
RT  "A transcribed gene in an intron of the human factor VIII gene.";
RL  Genomics 7:1-11(1990).
CC  -|- FUNCTION: NOT KNOWN. POSSIBLE HOUSEKEEPING ROLE.
CC  -|- TISSUE SPECIFICITY: PRODUCED ABUNDANTLY IN A WIDE VARIETY OF
CC  CELL TYPES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M34677; AAA35713.1; -.
DR  PIR; A34579; A34579.
DR  MIM; 305423; -.
SQ  SEQUENCE 364 AA; 38647 MW; 2C2B78DA0F2A820B CRC64;

```

Matches 39; Conservative 15; Mismatches 53; Indels 36; Gaps 6;

Qy 2 QAVAAAAEYAGLKVARROM-QDAAGRRHF-HASQCPRTSPVSTDSNMS----- 48
| |||| | | : | | | | : | : | :
Db 136 QPAAAAALCLELAALRDLDGQAAAAAGHFQRAAQQLQLPLAALQALGEAASQQLLARD 195

Qy 49 -----AVVIQKARPAKKQKHQPHGLRREAYADLLPPPPVPPPAIKSPVQSKAQLVPRV 103
| | : | : | | | | | | : | : | :
Db 196 YTGALAVFTMRQLAREHGSHP-----VQSLPPPPPPAPA-----RARGDARP- 238

Qy 104 MVPKLASIEARTDRSSDRKGGSY 126
: | | | : | : | :
Db 239 -TGRAASSELRLGGALSRLGAF 260

RESULT 6

CCAA_RABIT

ID CCAA_RABIT STANDARD; PRT; 2424 AA.
AC P27884; P27883;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE VOLTAGE-DEPENDENT P/Q-TYPE CALCIUM CHANNEL ALPHA-1A SUBUNIT (CALCIUM
CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE ISOFORM 4) (BRAIN CALCIUM CHANNEL
I) (BI).
GN CACNA1A OR CACNL1A4 OR CACH4 OR CACN3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=91187110; PubMed=1849233;
RA Mori Y., Friedrich T., Kim M.-S., Mikami A., Nakai J., Ruth P.,
RA Bosse E., Hofmann F., Flockerzi V., Furuichi T., Mikoshiba K.,
RA Imoto K., Tanabe T., Numa S.;
RT "Primary structure and functional expression from complementary DNA
of a brain calcium channel."
RL Nature 350:398-402(1991).
RN [2]
RP BETA-SUBUNIT BINDING DOMAIN, AND MUTAGENESIS.
RX MEDLINE=94150724; PubMed=7509046;
RA Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P.,
RA Campbell K.P.;
RT "Calcium channel beta-subunit binds to a conserved motif in the I-II
cytoplasmic linker of the alpha 1-subunit."
RL Nature 368:67-70(1994).
CC -1- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
GVIA).
CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: IN THE BRAIN, A SHORT ISOFORM BI-1/1A-1 AND
A LONG ISOFORM BI-2/1A-2 (SHOWN HERE), ARE PRODUCED BY ALTERNATIVE
SPLICING.
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN
PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
CURRENT IN CEREBELLAR GRANULE CELLS.
CC -1- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE

CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X57477; CAA40715.1; -.
DR EMBL; X57689; CAA40872.1; -.
DR EMBL; X57476; CAA40714.1; -.
DR EMBL; X57688; CAA40871.1; -.
DR INTERPRO; IPR000636; -.
DR INTERPRO; IPR002077; -.
DR PFAM; PF00520; ion_trans; 4.
DR PRINTS; P00167; CACHANNEL.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Alternative splicing.
FT REPEAT 85 363 I.
FT REPEAT 473 717 II.
FT REPEAT 1240 1523 III.
FT REPEAT 1560 1823 IV.
FT DOMAIN 1 98 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 99 117 S1 OF REPEAT I (POTENTIAL).
FT DOMAIN 118 135 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 136 155 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 156 167 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 168 185 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 186 190 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 191 209 S4 OF REPEAT I (POTENTIAL).
FT DOMAIN 210 228 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 229 248 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 249 335 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 336 360 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 361 487 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 488 506 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 507 521 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 522 541 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 542 549 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 550 568 S3 OF REPEAT II (POTENTIAL).
FT DOMAIN 569 578 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 579 597 S4 OF REPEAT II (POTENTIAL).
FT DOMAIN 598 616 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 617 636 S5 OF REPEAT II (POTENTIAL).
FT DOMAIN 637 689 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 690 714 S6 OF REPEAT II (POTENTIAL).
FT DOMAIN 715 1253 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1254 1272 S1 OF REPEAT III (POTENTIAL).
FT DOMAIN 1273 1288 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1289 1308 S2 OF REPEAT III (POTENTIAL).
FT DOMAIN 1309 1320 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1321 1339 S3 OF REPEAT III (POTENTIAL).
FT DOMAIN 1340 1350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1351 1369 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 1370 1388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1389 1408 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 1409 1495 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1496 1520 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1521 1575 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1576 1604 S1 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1605 1609 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1610 1629 S2 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1630 1637 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1638 1656 S3 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1657 1665 EXTRACELLULAR (POTENTIAL).

```

RESULT 8
SOS1_MOUSE
ID  SOS1_MOUSE          STANDARD;          PRT;  1319 AA.
AC  Q62245; Q62244;
DT  15-JUL-1999 (Rel. 38, Created)
DT  15-JUL-1999 (Rel. 38, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  SON OF SEVENLESS PROTEIN HOMOLOG 1 (SOS-1) (MSOS-1).
GN  SOS1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SWISS; TISSUE=EYE;
RX  MEDLINE=92335328; PubMed=1631150;
RA  Bowtell D., Fu P., Simon M., Senior P.;
RT  "Identification of murine homologues of the Drosophila son of
RT  sevenless gene: potential activators of ras.";
RL  Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
RN  [2]
RP  STRUCTURE BY NMR OF 415-548.
RX  MEDLINE=97360234; PubMed=9217262;
RA  Koshiba S., Kigawa T., Kim J.-H., Shirouzu M., Bowtell D.,
RA  Yokoyama S.;
RT  "The solution structure of the pleckstrin homology domain of mouse
RT  Son-of-sevenless 1 (mSosl).";
RL  J. Mol. Biol. 269:579-591(1997).
CC  -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC  SIMILARITY).

```

```

RESULT 10
FURI_BOVIN
ID FURI_BOVIN STANDARD; PRT; 797 AA.
AC Q28193;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FURIN PRECURSOR (EC 3.4.21.75) (PAIRED BASIC AMINO ACID RESIDUE
DE CLEAVING ENZYME) (PACE) (DIBASIC PROCESSING ENZYME) (TRANS GOLGI
DE NETWORK PROTEASE FURIN).
GN PACE OR FUR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;

```

RESULT 12
MYPH_CHICK

RESULT 13
EPS8_HUMAN
ID EPS8_HUMAN STANDARD; PRT; 822 AA.
AC Q12929:

Query Match 10.7%; Score 81.5; DB 1; Length 822;
Best Local Similarity 22.6%; Pred. No. 13;
Matches 44; Conservative 24; Mismatches 72; Indels 55; Gaps 8;

```

Qy 133 DGRQVTDLRTNPSDP 147
      | | | | |
Db 311 PEGEGLTLRAKPPPP 325

```


RESULT 15
ERF HUMAN

```

Qy      6 AAAEYAGLKVARQQMDAAGRRHFHASQCPRP-----TSPV-----STDNSMSAVVIQK 54
      | | | | : | | | : | | | : | | | :
Db    330 AFLHYPLGLVVPQQRPD-----KCPLPPMAPETPPVPSSASSSSSSSSSPFKKL 379
      | | | | : | | | : | | | : | | | :
Qy     55 ARPAAKQKHQGHRLREAYADD-----LPPPPVPPPAIKSPTVQSKAQLV 100
      | | : : : | | | : | | | | : | |
Db    380 QRPLGRQRRAAGEKVAANAADKSGGSAGGLAEGAGALAPPPPPP-----QIKV 427
      | | : : : | | | : | | | | : | |
Qy    101 RPVMVPKIASIEARTDRSSDRKGSYKGREALDGRQVTLRLTNPSDPR 148
      | : : | | | | : | | | : | | |
Db    428 EPISEGESEVLEV-TDISDE-----DEEDGEVFKTPRAPPAAPK 465
      | | : : : | | | : | | | : | | |

```

Search completed: January 22, 2001, 12:29:56
Job time: 1297 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2001, 12:54:12 ; Search time 559.88 Seconds
(without alignments)
30.983 Million cell updates/sec

Title: US-09-540-245A-20
Perfect score: 761
Sequence: 1 AQAVAAAAEYAGLKVARROM.....REALDGRQVTDLRTPNSDPR 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	761	100.0	1612	11	O89026	O89026 mus musculus
2	719	94.5	1651	11	O55005	O55005 rattus norvegicus
3	691	90.8	1651	4	Q9Y6N7	Q9Y6N7 homo sapiens
4	101	13.3	1344	11	Q9Z2I4	Q9Z2I4 mus musculus
5	99	13.0	739	5	O45408	O45408 caenorhabditis
6	97.5	12.8	413	12	P89474	P89474 herpes simplex
7	97.5	12.8	1262	5	Q9VWC2	Q9VWC2 drosophila
8	91	12.0	825	5	Q23484	Q23484 caenorhabditis
9	91	12.0	1017	10	Q9SS68	Q9SS68 arabidopsis
10	91	12.0	3080	5	Q9VRY3	Q9VRY3 drosophila
11	90.5	11.9	410	13	Q9IA21	Q9IA21 heterodontomys
12	90	11.8	1212	10	Q9LGT8	Q9LGT8 oryza sativa
13	89	11.7	569	4	Q9UGJ0	Q9UGJ0 homo sapiens
14	89	11.7	975	5	Q9VLA7	Q9VLA7 drosophila
15	89	11.7	2936	5	Q9NKP7	Q9NKP7 leishmania
16	88	11.6	811	5	Q9V8T0	Q9V8T0 drosophila
17	88	11.6	831	5	Q9NFL3	Q9NFL3 drosophila
18	88	11.6	980	11	P97837	P97837 rattus norvegicus
19	87.5	11.5	621	10	Q9LW19	Q9LW19 arabidopsis

20	87	11.4	764	5	Q24708	Q24708 drosophila
21	86.5	11.4	296	10	Q40225	Q40225 lillium longifolium
22	86.5	11.4	891	5	Q93763	Q93763 caenorhabditis
23	86	11.3	1650	5	Q9NES4	Q9NES4 caenorhabditis
24	85.5	11.2	604	10	Q9LW37	Q9LW37 arabidopsis
25	85.5	11.2	662	5	Q21536	Q21536 caenorhabditis
26	85	11.2	300	10	Q9LY35	Q9LY35 arabidopsis
27	85	11.2	698	5	O44447	O44447 caenorhabditis
28	85	11.2	716	5	Q9U2A6	Q9U2A6 caenorhabditis
29	85	11.2	1541	5	O15837	O15837 leishmania
30	84.5	11.1	2089	4	Q14676	Q14676 homo sapiens
31	84.5	11.1	2099	4	Q9Y2W9	Q9Y2W9 homo sapiens
32	84.5	11.1	2099	4	Q9UNU8	Q9UNU8 homo sapiens
33	84.5	11.1	3201	5	Q9W0U2	Q9W0U2 drosophila
34	84.5	11.1	4880	11	Q9JLT1	Q9JLT1 rattus norvegicus
35	84.5	11.1	5085	11	Q9JKS6	Q9JKS6 rattus norvegicus
36	84	11.0	193	5	O76215	O76215 neospora caninum
37	84	11.0	412	12	O10415	O10415 helicoverpa
38	84	11.0	754	4	Q9NUP6	Q9NUP6 homo sapiens
39	83.5	11.0	282	10	Q40509	Q40509 nicotiana glauca
40	83.5	11.0	365	5	Q9XUP5	Q9XUP5 caenorhabditis
41	83.5	11.0	1180	5	Q9VRM2	Q9VRM2 drosophila
42	83.5	11.0	3232	5	Q9VFF8	Q9VFF8 drosophila
43	83	10.9	495	4	Q9NV83	Q9NV83 homo sapiens
44	83	10.9	520	4	Q9P2K3	Q9P2K3 homo sapiens
45	83	10.9	787	3	O94096	O94096 pneumocystis

ALIGNMENTS

RESULT 1
O89026
ID O89026 PRELIMINARY; PRT; 1612 AA.
AC O89026;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DUTTL PROTEIN.
GN ROBO1 OR DUTTL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Wu M.C., Lowe N., Fordham R., Rabbitts P.;
RT "The mouse homologue of human DUTTL/H-robo1 gene: protein sequence and
RT chromosomal location."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17793; CAA76850.1; -.
DR HSSP; P56276; 1TLK.
DR MGD; MGI:1274781; Robo1.
DR INTERPRO; IPR001777; -.
DR INTERPRO; IPR003006; -.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00047; ig; 5.
SQ SEQUENCE 1612 AA; 176406 MW; 5F2988C544796B4B CRC64;

Query Match 100.0%; Score 761; DB 11; Length 1612;
Best Local Similarity 100.0%; Pred. No. 2.5e-65;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQAVAAAAEYAGLKVARROMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAK 60
Db 1355 AQAVAAAAEYAGLKVARROMQDAAGRRHFHASQCPRPTSPVSTDSNMSAVVIQKARPAK 1424
Qy 61 QKHQPGHLRREAYADLPPPPVPPPAIKSPTVQSKAQLEVRPVMVPKLASIEARTDRSSD 120
Db 1425 QKHQPGHLRREAYADLPPPPVPPPAIKSPTVQSKAQLEVRPVMVPKLASIEARTDRSSD 1484

RESULT 5
045408


```

RESULT 9
Q9SS68
ID Q9SS68 PRELIMINARY; PRT; 1017 AA.
AC Q9SS68;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PUTATIVE PHOSPHORIBOSYLANTHRANILATE TRANSFERASE.
GN T12J13.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12J13 genomic sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009327; AAF03465.1; -.
DR INTERPRO; IPR006008; -.
DR PFAM; PF00168; C2; 3.
DR PROSITE; PS50004; C2 DOMAIN 2; 3.

```

```
Qy 105 -----VPKLASIEARTDRSSDRKGGSYKGREALDGRQVTDL 140
      :|:|: | | : | | | : :|
Db 139 ESRONAKOKTSSSSSVESSAGEKSPPGPASKRARTAYTSAOLVEL 183
```

QY 25 GRRHFHASOCPRPTSPVSTD-SNMSAVVIOKARPAKKOKHOPGHLREAYADDLPP---- 79

Query Match 11.7%; Score 89; DB 5; Length 975;
Best Local Similarity 24.5%; Pred. No. 1.4;
Matches 34; Conservative 22; Mismatches 67; Indels 16; Gaps 4;

Search completed: January 22, 2001, 12:54:19
Job time: 2060 sec

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.